

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: March 4, 2004, 18:36:21 ; Search time 147.5 Seconds
(without alignments)
12757.738 Million cell updates/sec

Title: US-10-007-270-1
Perfect score: 5875
Sequence: 1 taacaacgaaggttatctct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xl
-Q/cn2_1/USPTO.spool_p/US10007270/runat 04032004 160715 4884/app_query.fasta_1.3527
-DB=A Geneseg 29Jan04 -QMW=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10007270 @CN 1 1 117 @runat 04032004 160715 4884 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WREN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseg 29Jan04.*
1: Genesegp1980s.*
2: Genesegp1990s.*
3: Genesegp2000s.*
4: Genesegp2001s.*
5: Genesegp2002s.*
6: Genesegp2003s.*
7: Genesegp2003bs.*
8: Genesegp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|------------|--------------------|
| 1 | 4073 | 69.3 | 797 | 6 ADAL4841 | Adal4841 Human int |
| 2 | 4073 | 69.3 | 797 | 7 ABR42342 | AbR42342 Human int |
| 3 | 3862 | 65.7 | 771 | 3 AAY57089 | Aay57089 Human int |
| 4 | 3862 | 65.7 | 771 | 3 AAY93336 | Aay93336 A human i |
| 5 | 3862 | 65.7 | 771 | 6 ADAL4867 | Adal4867 Human var |
| 6 | 3862 | 65.7 | 771 | 7 ABR42354 | AbR42354 Human int |
| 7 | 3611 | 61.5 | 719 | 6 ADAL4843 | Adal4843 Human int |
| 8 | 3611 | 61.5 | 719 | 7 ABR42343 | AbR42343 Human int |
| 9 | 2504.5 | 42.6 | 798 | 6 ADAL4848 | Adal4848 Mouse int |
| 10 | 2504.5 | 42.6 | 798 | 7 ABR42345 | AbR42345 Mouse int |

| | | | | | |
|----|--------|------|------|------------|--------------------|
| 11 | 2171 | 37.0 | 709 | 3 AAY93338 | Aay93338 A murine |
| 12 | 1298.5 | 22.1 | 466 | 6 ADAL4850 | Adal4850 Mouse int |
| 13 | 1298.5 | 22.1 | 466 | 7 ABR42346 | AbR42346 Mouse int |
| 14 | 1038 | 17.7 | 198 | 6 ADAL4845 | Adal4845 Human int |
| 15 | 1038 | 17.7 | 198 | 7 ABR42344 | AbR42344 Human int |
| 16 | 918 | 15.6 | 1241 | 3 AAY93337 | Aay93337 A human i |
| 17 | 917 | 15.6 | 1241 | 6 ADAL4856 | Adal4856 Human int |
| 18 | 917 | 15.6 | 1241 | 7 ABR42349 | AbR42349 Human int |
| 19 | 841 | 14.3 | 185 | 3 AAY93335 | Aay93335 An interp |
| 20 | 841 | 14.3 | 185 | 6 ADAL4854 | Adal4854 Monkey in |
| 21 | 841 | 14.3 | 185 | 7 ABR42348 | AbR42348 Monkey in |
| 22 | 820.5 | 14.0 | 1069 | 3 AAY93339 | Aay93339 A murine |
| 23 | 820.5 | 14.0 | 1069 | 6 ADAL4863 | Adal4863 Mouse int |
| 24 | 820.5 | 14.0 | 1069 | 7 ABR42352 | AbR42352 Mouse int |
| 25 | 420.5 | 7.2 | 432 | 6 ADAL4858 | Adal4858 Human int |
| 26 | 420.5 | 7.2 | 432 | 7 ABR42350 | AbR42350 Human int |
| 27 | 196 | 3.3 | 577 | 3 AAY71030 | Aay71030 Ubiquitin |
| 28 | 189.5 | 3.2 | 1255 | 5 ABP56040 | Abp56040 MUC1 rece |
| 29 | 189.5 | 3.2 | 1255 | 6 ABR82568 | AbR82568 MUC1/REP |
| 30 | 189.5 | 3.2 | 1255 | 6 ABR47537 | AbR47537 Breast ca |
| 31 | 189.5 | 3.2 | 1255 | 6 ABR92124 | AbR92124 Human cer |
| 32 | 189.5 | 3.2 | 1255 | 7 ADD45111 | Add45111 Human Pro |
| 33 | 189.5 | 3.2 | 1255 | 7 ADE54622 | Ades4622 Human Pro |
| 34 | 188.5 | 3.2 | 495 | 4 AAU00539 | Aau00539 Human MUC |
| 35 | 188.5 | 3.2 | 515 | 5 ABG96378 | Abg96378 Human ova |
| 36 | 188.5 | 3.2 | 515 | 6 ADA50565 | Ada50565 Human muc |
| 37 | 188.5 | 3.2 | 515 | 6 ABR92123 | AbR92123 Human cer |
| 38 | 188.5 | 3.2 | 515 | 6 AAE37797 | Aae37797 Human muc |
| 39 | 185.5 | 3.2 | 508 | 2 AAH77233 | Aah77233 MiniMUC1 |
| 40 | 183.5 | 3.1 | 455 | 3 AAH71024 | Aay71024 Human MUC |
| 41 | 183.5 | 3.1 | 475 | 4 AAU00573 | Aau00573 Human MUC |
| 42 | 183.5 | 3.1 | 475 | 5 ABE77476 | Abb77476 Human MUC |
| 43 | 183.5 | 3.1 | 475 | 6 ADA50567 | Ada50567 Mucin 1 (|
| 44 | 183.5 | 3.1 | 475 | 6 AAE37800 | Aae37800 Human muc |
| 45 | 183.5 | 3.1 | 475 | 7 ADD14120 | Add14120 Human src |

ALIGNMENTS

RESULT 1

ADAL4841

ID ADAL4841 standard; protein; 797 AA.

XX AC ADAL4841;

XX DT 06-NOV-2003 (first entry)

XX DE Human interphotoreceptor matrix component, IPMC, 150 isoform A.

XX KW human; IPMC 150 isoform A; gene therapy;

XX KW interphotoreceptor matrix component; IPMC; ocular disorder;

XX KW macular degeneration; photoreceptor death; retinal detachment.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Peptide 1..20

XX FT Protein /note= "Signal sequence"

XX FT Protein 21..797

XX FT Misc-difference /note= "Mature IPMC 150 isoform A"

XX FT /note= "187..188"

XX FT /note= "Encoded by ACACGATCTTCATTTCAAGACTTGGCGAGTATTCTA

XX FT AGAARACCTTCAGAGCAATTCAGAT"

XX US2002160954-A.

XX 31-OCT-2002.

XX 08-NOV-2001; 2001US-00007270.

XX 29-OCT-1998; 98US-00183972.

XX 29-OCT-1999; 99US-00430195.

XX

PA (IOWA) UNIV IOWA RES FOUND.
 XX Hageman GS, Kuehn MH;
 XX WPI; 2003-238235/23.
 DR N-PSDB; ADA14840.
 XX
 XX New isolated or recombinant interphotoreceptor matrix component
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,
 PT treating or prognosticating ocular disorders, e.g. macular degeneration
 PT or retinal detachment.
 XX
 PS Claim 8; Page 31-33; 76pp; English.
 XX

CC The invention relates to an isolated or recombinant interphotoreceptor
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)
 CC gene operatively linked to the IPMC polynucleotide. The IPMC
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,
 CC preventing, treating or prognosticating ocular disorders, e.g. macular
 CC degeneration, photoreceptor death or retinal detachment. They are also
 CC useful for identifying a compound capable of modulating IPMC gene
 CC expression in a cell. The present sequence represents the amino acid
 CC sequence of human interphotoreceptor matrix component, IPMC, 150 isoform
 CC A.

SQ Sequence 797 AA;

Alignment Scores:
 Pred. No.: 0 Length: 797
 Score: 4073.00 Matches: 796
 Percent Similarity: 97.31% Conservative: 1
 Best Local Similarity: 97.19% Mismatches: 0
 Query Match: 69.33% Indels: 22
 DB: 6 Gaps: 1

US-10-007-270-1 (1-3330) x ADA14841 (1-797)

QY 131 ATGTATTGGAACTAGAGAGCTATTGTTGTTTGGATTTTCTCAAGTTCAGGA 190
 DB 1 MetTyLeuGluThrArgArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCTCCATTACATATACCATCTGAACTAAAGACATAGACATCCCA 250
 DB 21 ThrIysAspIleSerIleAenIleTyHisSerGluThrLysAspIleAspAsnProPro 40
 QY 251 AGAATGAACAACTGAAGTATGAAAAAATGTCAAAATGTCAACTATGAGCAATA 310
 DB 41 ArgAsnGluThrThrGluSerThrGluLysMetTyLysMetSerThrMetArgArgIle 60
 QY 311 TTCGATTGGCAAGCATCGAACAAAGATCCGATTTTCCGACGGGGTTAAAGTC 370
 DB 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
 QY 371 TGTCACAGGAATCCATGAACAGATTTTACAGAGTCTTCAAGCTTATTATAGATTGAGA 430
 DB 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyArgLeuArg 100
 QY 431 GTCTGTCAGGAGCAGTATCGGAAGCATATCGGATCTTCTCGATCGCATCCCTGACACA 490
 DB 101 ValCysGlnGluAlaValTrpGluAlaTyArgIlePheLeuAspArgIleProAspThr 120
 QY 491 GGGGAATATCAGGACTGGGTACGATCTGCCAGCAGGAGCCTTCTGCTTTCGATTCAT 550
 DB 121 GlyGluTyGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140
 QY 551 GGAATAATCTCAGCAATCCAGGAGCAGCTGGATCTTCTCCAGCAGAGATATAAAGAG 610
 DB 141 GlyLysAsnPheSerAsnSerGlnGlnHisLeuAspLeuLeuGlnGlnHisLeuGln 160
 QY 611 AGAAGTTTCCCTGACAGAAAGATGAATATCTGCAAGAGACATTTGGAGAGCCTGGT 670
 DB 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGlnGluProGly 180

QY 671 GAAACCATTTGTCATTTCAACAGCAATCTACATTTCAAGAGCTTGGGACAGTATTCTAAGAA 730
 DB 181 GluThrIleValIleSer--Thr----- 187
 QY 731 AACCTCAGAGAGCAAAATTCAGAGTGTGGCAACGTCTCATTGGGCTTTCCCTCTCA 790
 DB 188 -----AspValAlaAenValSerLeuGlyProPheProLeu 200
 QY 791 CTCTGATGACACCTCTCTCAATGAATTTCTCGATATATACCTCAACAGCAGCAGATGC 850
 DB 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 220
 QY 851 CTACACAGAGAGAGAAACAGATTCTGCTGTGTGGAGGAGCAGAGGTGGAGTCAAGCG 910
 DB 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGluGluGluGluGluGlu 240
 QY 911 TCTCTCTGTAAACACAGAGTTCAGAGCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970
 DB 240 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTy-TyrG 260
 QY 971 AGAGCTACGAGGAAAGTCCCACTTCAGATGCCAAAGATATTTAAGAAACTTCCAGAT 1030
 DB 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLeuProGlyP 280
 QY 1031 TCAGAAATCCATGTTAGATTAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1090
 DB 280 helyLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 300
 QY 1091 CAGAGATCAACTTACGGCCATCTTTAAGAGACACAGTGCAGAGAGCAAGAGAGAGAGAG 1150
 DB 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAla 320
 QY 1151 GTGACCTCTGCTTTTGTGATTCACAAATTTGAAGTGCAGAGAGTCTATCATGCAACA 1210
 DB 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyHisGlyThrM 340
 QY 1211 TCGAG 1270
 DB 340 eLysGluLysGlnProGluIleTyLeuThrAlaThrAspLeuLysArgLeuLys 360
 QY 1271 GCAAGACCTAG 1330
 DB 360 erLysAlaLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 380
 QY 1331 TTGCTGATCATCGACGCTTTGGTCTGCACACCAATCAGAGCTGCCCACTTTTTCG 1390
 DB 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPhe 400
 QY 1391 CTGTTTATACAG 1450
 DB 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeu 420
 QY 1451 AGACAGTGCAG 1510
 DB 420 luthrValAlaAspGlyAlaGluHisGlyLeuProAspThrSerIlePheProAlaMet 440
 QY 1511 CCTTACCTCTCTGTCAGAGAGTCCACCTTTTATGTCATCAAGCATCTTCTCTCTCA 1570
 DB 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeu 460
 QY 1571 CTGATCAAGGACACAGAGATACAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1630
 DB 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeu 480
 QY 1631 CCATCCCCACAGTGAATTTTTCGAATTCAGCAACTGCTCTGGGAATTTTCATCTCAC 1690
 DB 480 hrIleProThrSerAspTySerAlaIleSerGlnLeuAlaGlyLysSerHisProp 500
 QY 1691 CTGCATCTTCAGATGACAGCCCATCAAGTGCAGGTGGCGAGATATGGTCAGACACCTAG 1750
 DB 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGluAspMetValArgHisLeu 520

QY 1751 ATGAATGATCTGTCTGACATCTCTGCCCCATCTGAGTACCAAGCTCAGGCAATATG 1810
 Db |||||
 QY 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540
 Db |||||
 QY 1811 TTCTGTGTCGAGATCATCTTCTGGAGATACCACTCTCTCTCTCAGCTTACAGTATATCA 1870
 Db |||||
 QY 540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGluTyrIleT 560
 Db |||||
 QY 1871 CCATAGTCTTATGACCATGCCCCAGGCGGAGAGCTGTAGTGTCTTCTCAGTCTGC 1930
 Db |||||
 QY 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580
 Db |||||
 QY 1931 GTGTGTTAAATGCGCTTCTCCAAACGACCTGTTCAAAGAGCTCTCTGAGTACCGAG 1990
 Db |||||
 QY 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArg 600
 Db |||||
 QY 1991 CTTCTGGAGCAACATTCACACAGCTGTCTTCCATCTACGATCCATCTTACAGAT 2050
 Db |||||
 QY 600 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620
 Db |||||
 QY 2051 TTAAGCAACTGAAATACCTTAACCTCACAAGCTGTGCGAGGGGCTTGGAGGATT 2110
 Db |||||
 QY 620 heLysGlnLeuGluLeuLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
 Db |||||
 QY 2111 AGTTTGTCTAGTCTGCGGTATTAACCTCACAAGCTGTGCGAGGGGCTTGGAGGATT 2170
 Db |||||
 QY 640 ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 660
 Db |||||
 QY 2171 TTCCTTCTGTCGAGCCCAACCACTCTGGAATAGACAGTCTCTCTCAACATG 2230
 Db |||||
 QY 660 heArgSerAlaAlaGlnGlnLeuHisLeuGluLeuAspSeryrSerLeuAsnIleG 680
 Db |||||
 QY 2231 AACGAGCTGATCAAGCAGATCCCTGCAAGTCTCTGCGCTGCGCGAATTCGCCAATGTG 2290
 Db |||||
 QY 680 luProAlaAspGlnAlaAspProCysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
 Db |||||
 QY 2291 TTAAGAACGACGCTGAGAAAGCGAGTGTCTGCTCAAAACAGGATATCACAGCCAGG 2350
 Db |||||
 QY 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyrAspSerGlnG 720
 Db |||||
 QY 2351 GGAGCTGTCGAGCGTGTGACACGAGGCTCTGTGCTGCTGCAAAAGGATGCGAGTCC 2410
 Db |||||
 QY 720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 740
 Db |||||
 QY 2411 TCAGGAAAGGAGCTCCATGCGAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAGT 2470
 Db |||||
 QY 740 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLys 760
 Db |||||
 QY 2471 CTAGTGTAAAGAGTTCGAAATCAACAAATACAGGTAATACAGTAAAGAAATCTG 2530
 Db |||||
 QY 760 hrSerValLysPheGlnAsnGlnAsnLysValIleSerLysArgAsnSerG 780
 Db |||||
 QY 2531 AATTACTGACGTAGATATGAAATTTAACATCAAGATGGAAGGAAT 2583
 Db |||||
 QY 780 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 797

RESULT 2
 ABR42342
 ID ABR42342 standard; protein; 797 AA.
 AC ABR42342;
 XX
 XX
 DT 11-AUG-2003 (first entry)
 DE Human interphotoreceptor matrix IPM 150, isoform A.
 XX
 KW Human; interphotoreceptor matrix; IPM 150; IPMC; receptor;
 XX opthalmological; gene therapy.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1. .20

Protein /label= Signal_peptide
 21..797
 Domain /label= IPM 150
 71..88
 Domain /note= "N-terminal domain"
 95..115
 Domain /note= "conserved domain"
 187..188
 Misc-difference /note= "EGF-like domain"
 688..731
 Domain /note= "EGF-like domain"
 WO2003039346-A2.
 15-MAY-2003.
 08-NOV-2002; 2002WO-US036090.
 08-NOV-2001; 2001US-00077270.
 (IOWA) UNIV IOWA RES FOUND.
 Hageman GS, Kuehn MH;
 WPI; 2003-441440/41.
 N-PSDB; ACC57946.
 New interphotoreceptor matrix proteins and polynucleotides, useful for treating or preventing photoreceptor death or retinal detachment, or for treating ocular disorders.
 Claim 6; Page 78; 105pp; English.
 The present sequence is the protein sequence of isoform A of novel human interphotoreceptor matrix 150 (IPM 150), a member of the newly identified interphotoreceptor matrix component (IPMC) gene family. The IPM 150 gene is located on chromosome 6q13-q15, a region that also contains loci for progressive bifocal choriorretinal atrophy, autosomal dominant Stargardt's-like macular dystrophy, North Carolina macular dystrophy and Salla disease. Members of the IPMC gene family have been identified in humans, monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2 Subfamilies are designated IPM 150 (or IPM1) and IPM 200 (or IPM2). The invention provides IPM 150 and IPM 200 polynucleotides and polypeptides, antibodies that specifically bind the polypeptides, and vectors comprising the polynucleotides. A claimed method of treating or preventing photoreceptor death or retinal detachment involves administering an IPMC polynucleotide, polypeptide or antibody. Also claimed is a method for identifying a compound capable of modulating IPMC gene expression
 SQ Sequence 797 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 797
 Score: 4073.00 Matches: 796
 Percent Similarity: 97.31% Conservative: 1
 Best Local Similarity: 97.19% Mismatches: 0
 Query Match: 69.33% Indels: 22
 DB: 7 Gaps: 1
 US-10-007-270-1 (1-3330) x ABR42342 (1-797)
 QY 131 ATGTATTGGAACTAGAGAGCTATTGTTTTGGATTTTCCCAAGTTCAAGA 190
 Db |||||
 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAAGATATCTCCATTAACATATACCATCTGAACTAAAGACATAGCAATCCCCCA 250
 Db |||||
 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAsnProPro 40
 QY 251 AGAAATGAAACAACTGAAAGTACTGAAATAATGTCAAAATGTCAACTATGAGAGATA 310
 Db |||||

Db 41 ArgAsnGluThrThrGluSerThrGluLysMetThrLysMetSerThrMetArgArgIle 60
 QY 311 TTCCGATTGGCAAGACATCGAACAAAGATCCGCAATTTTCCCAACGGGGTAAAGTC 370
 Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValIysVal 80
 QY 371 TGTCCACAGGAATCCATGAACAGATTTTACAGATCTTCAAGCTTATATAGATTGAGA 430
 Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaIleLysArgLeuArg 100
 QY 431 GTGTCTCAGGAGCAGTATGGAGCATATCGGATCTTTCGATCGCATCCCTGCACACA 490
 Db 101 ValCysGlnGluAlaValTrpGluAlaIleArgIlePheLeuAspArgIleProAspThr 120
 QY 491 GGGGAATATCAGGACTGGCTCAGCATCTGCCAGCAGGAGACTTCTGCTCTTTGCATT 550
 Db 121 GlyGluTrpGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140
 QY 551 SGAAAAAATCTCAGCAATTCGAGGACCTGGATCTTCTCCACAGCAGAGATAAAGACAG 610
 Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleLysGln 160
 QY 611 AGAAGTTTCCCTCAGCAGAAAGATGAATATCTGCAGAGAGACATTTGGAGAGCCTGGT 670
 Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluProGly 180
 QY 671 GAAACCATCTCATTTTCAACAGCAATCTACATTTCAAGACTTGGCAGATTCTTAAGAA 730
 Db 181 GluThrIleValIleSer--Thr----- 187
 QY 731 AACCTCAGAGAGCAAAATTCAGATGTTGCCAAGCTCTCACTTGGGCTTTCCCTCTCA 790
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeu 200
 QY 791 CTCCTGATGACACCTCTCAATGAATTCGATATACATCAACGACACCAAGATGC 850
 Db 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 220
 QY 851 CTACACAGAAAGAAACAGAAATTCGCTGTGTTGGAGGAGCAGAGGCTGAGCTCAGG 910
 Db 220 rothrThrGluArgGluThrGluPheAlaValLeuGluGluGluArgValGluLeuSerV 240
 QY 911 TCTCTCTGTAAACAGAGTTCAGGAGAGCTGCTGACTCCAGCTCCCACTATATACC 970
 Db 240 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG 260
 QY 971 AGGAGCTAGCAGAACTCCCACTTCAGATGCAAGAGATATTAAAGAACTTCCAGGAT 1030
 Db 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP 280
 QY 1031 TCAAAAAATCCATGTGTAGATTTAGACCAAGAAAGAAAGAAAGATGCTCAAGTCCA 1090
 Db 280 helLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 300
 QY 1091 CAGAGATCACTTACGGCCATCTTAAAGACACAGTGCAGAGCAAAAGCCCTGCAA 1150
 Db 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlaS 320
 QY 1151 GTGACCTCTGCTTTTGTATCCCAAAATTCGAAGTGAAGAGTCTATCATGAACCA 1210
 Db 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 340
 QY 1211 TGGAGGAGCAGAGCAACAGAAATCTATCTCAGCTACAGACTTCAAAAGGCTGATCA 1270
 Db 340 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360
 QY 1271 GCAAGCACTAGAGGAGAGAACATCTTTGGATGTGGGACAAATTCAGTTCACTGATGAA 1330
 Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380
 QY 1331 TTGCTGGATCACTGCCAGCTTGTGCTCTGACACCCATCAGAGCTGCCACATCTTTTG 1390
 Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400

QY 1391 CTGTTTATACAGAGGATGCTACTTTGAGTCCAGAACTTCTCTCTGTTGAAACCCAGCTTG 1450
 Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420
 QY 1451 AGACAGTGCAGGAGCAGACATGTGCTACCTGACACACTTCTTGGTCTCCACTGCTATGG 1510
 Db 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440
 QY 1511 CCTTACCTCTCTCAGAGCTCCACCTTTCTTTATGCTCAAGCATCTCTCTCTCTCA 1570
 Db 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeu 460
 QY 1571 CTGATCAGGACACACAGATACAAATGGCCACTGACAGACAAATGCTAGTACACAGGCTCA 1630
 Db 460 hrAspGlnGlyThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeu 480
 QY 1631 CCATCCCAACAGTGTATTCTGCAATCAGCAACCTGGCTCTGGGAAATTTCACTCCAC 1690
 Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProP 500
 QY 1691 CTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGGAAAGATATGCTCAGACACCTAG 1750
 Db 500 roAlaSerSerAspAspSerArgSerAlaGlyGlyGluAspMetValArgHisLeuA 520
 QY 1751 ATGAAATGGATCTGTCGACACTCTCTGCCCATCTGAGGTACAGAGCTCAGCGAATATG 1810
 Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540
 QY 1811 TTTCTGTCCAGATCAATTTCTGGAGATACCACTCTGCTCTCAGCTTTTACAGTATATCA 1870
 Db 540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIle 560
 QY 1871 CCATCTAGTCTATGACCATTCGCCCAAGGCCGAGAGCTGGTAGTGTCTTCTCAGTCTGC 1930
 Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580
 QY 1931 GTGTGTCTAACATGCTCTCCCAAGCTGTTTCAACAGAGCTCTCTGGAGTACCGAG 1990
 Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArg 600
 QY 1991 CTCTGGAGCAACAATTCACACAGCTCTGCTGTTTCCATATCTACGATCCCAATCTTACAGGT 2050
 Db 600 laLeuGluGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620
 QY 2051 TTAAGCAACTTCAATCTTAACCTCAGAAACGGAGTGTGATTTGTAATGACAAATCA 2110
 Db 620 helYsGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
 QY 2111 AGTTTCTAAGTCTGTGCTGCTATACCTTACCTCAGAGCTGTGACGGGTCTTGGAGATT 2170
 Db 640 ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660
 QY 2171 TTGTTCTGTGCGAGCCCAACACTCCATCTGGAATATAGACAGCTACTCTCTCAACATG 2230
 Db 660 heArgSerAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680
 QY 2231 AACAGCTGATCAGCAGATCTCTGCAAGTCTTGGCTTGGCGCAATTTGGCCAATGTG 2290
 Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
 QY 2291 TAAAGACAGCAAGCTGAGGAGCGAGTGTGCTGCAACACAGGATATGACAGCCAGG 2350
 Db 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720
 QY 2351 GGAGCTGGAGCGTCTGGAAACAGGCTCTGTGGGCTTGGCACAAGAGATCCAGGTC 2410
 Db 720 lysLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 740
 QY 2411 TCCAGGGAAGGAGCTCCATGAGGTGCTGAGATCACTCTGAAAATCAAGATACAAAA 2470
 Db 740 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLysT 760

US-10-007-270-1 (1-3330) x AAY57089 (1-771)

Db 320 erAspLeuLeuSerPheAsnLysIleGluSerGluGluValTyHieGlyThrm 340
QY 1211 TGGAGGAGGACAGCAACAGAAATCTATCTACAGCTACAGACCTCAAAAGGCTCATCA 1270
Db 340 etGluGluAspGlnProGluIleTyLeuThraLathrAspLeuLysArgueLies 360
QY 1271 GCAAGCACTAGAGAGAGAACATCTTTGGATGTGGGGAATTCAGTTCATGTAGAAA 1330
Db 360 erLysAlaLeuGluGluGlnSerLeuLeuAspValGlyThrIleGlnPheThrAspGluI 380
QY 1331 TTCTGGATCTCTCCAGCCTTTGGTCTGACACCCCAATCAGAGCTGCCACATCTTTTGG 1390
Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400
QY 1391 CTGTTTATACAGAGGATCTACTTTGAGTCCAGATCTCTCTCTGTTGACCCCAAGCTTG 1450
Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProProValGluProGlnLeuG 420
QY 1451 AGACAGTGGAGGAGCAGAGCATGTCTACCTGACACTCTCTGCTCTCCACCTGCTATGG 1510
Db 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440
QY 1511 CCTCTACTCTCTGTCAGAGCTCCACCTTTCTTTATGSCATCAAGCATCTCTCTCTGA 1570
Db 440 laSerThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT 460
QY 1571 CTGATCAAGGCACACAGATACATGCTCCACTGACAGACATGCTAGTACAGGGCTCA 1630
Db 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480
QY 1631 CCATCCCAACAGATATTCTGCAATCAGCCACTGGCTCTGGCAATTCACATCCAC 1690
Db 480 hrIleProThrSerAspTySerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProP 500
QY 1691 CTGCATCTTCAGATCAGCCGATCAAGTGCAGTGGCGAGATATGTTCAGACACTAG 1750
Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520
QY 1751 ATGAAATGATCTGTCTGACACTCTCCGCTCCCTGAGTACAGAGCTCTCGAGTACGAG 1810
Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyV 540
QY 1811 TTCTGTCCAGATCATTTCTGGAGATACCACTCTCTGCTCAGCTGTGTCTCTCAGTCTGC 1930
Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 580
QY 1931 GTGTGCTAAATGGCTCTCTCAAGCAGCTGTTCACAGAGCTCTCTGGAGTACGAG 1990
Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyArgA 600
QY 1991 CTCTGGAGCAATTCACAGCTGTCTGCTTCCATCTACGATCCAACTCTTACAGAT 2050
Db 600 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyLeuArgSerAsnLeuThrGlyP 620
QY 2051 TTAAGCAACTTGAATCTACTTCAAGAACCGAGTGTGTGATTTGCAATAGCAAAATGA 2110
Db 620 heLysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
QY 2111 AGTTTGTCTGTGTGCTGCTTAACTTCAAGAGCTGTGCGAGGCTCTGGAGGAT 2170
Db 640 ysPheAlaLysSerValProTyAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660
QY 2171 TTCGTCTGTGAGCCCAACCACTCCATCTGGAATAGACAGCTACTCTCTCAACATG 2230
Db 660 heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTySerLeuAsnIleG 680
QY 2231 AACAGCTGTATCAGAGNATCCGCAAGTCTCTGGCTGGCGGAATTTGCCAATGTG 2290
Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700

QY 2291 TAAAGAACCAACGAGTGGAGAGCGAGTCTGCTCAAAACCCAGGATATGACAGCCAGG 2350
Db 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyAspSerGlnG 720
QY 2351 GGAGCTGACCGTCTGGACACAGGCTCTCTGTGGCCCTGGCACAAGGAATGCGAGGTCC 2410
Db 720 lySerLeuAspGlyLeuGluProGlyLeuCysGly-LeuAlaGlnArgAsnAlaArgSer 739
QY 2411 TCCAGGAAAGGAGCTCCATGTCAGTTCGCCAGATCACTCTGAAATCAGCATACAAA 2470
Db 740 SerArgGluArgGluLeuHisAla-ValProAspHisSerGluAsnGlnAlaTyLysT 759
QY 2471 CTAGTCTTAAAGATTCACAAATCAACAAATACCAAGG 2509
Db 759 hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771
RESULT 4
AAV93336
ID AAV93336 standard; protein; 771 AA.
XX AC AAV93336;
XX DT 04-SEP-2000 (first entry)
XX DE A human interphotoreceptor matrix proteoglycan (IPM150).
XX KW Interphotoreceptor matrix; IPM; proteoglycan; IPM150; IPMC; IPM200;
XX KW chromosome 6q13-q15; ocular disease; retinal detachment;
XX KW choriorretinal degeneration; retinal degeneration; cone degeneration;
XX KW age related macular degeneration; photoreceptor degeneration;
XX KW retinal pigment epithelium degeneration; mucopolysaccharidosis;
XX KW rod- cone dystrophy; cone-rod dystrophy.
XX OS Homo sapiens.
FH Key
FT Domain 42..215 Location/Qualifiers
FT Modified-site 42 /note= "domain A"
FT Modified-site 143 /note= "potential glycosylation site"
FT Modified-site 191 /note= "potential glycosylation site"
FT Modified-site 215 /note= "potential glycosylation site"
FT Domain 221..565 /note= "potential glycosylation site"
FT Modified-site 221 /note= "domain B"
FT Modified-site 222 /note= "potential O-linked glycosylation site"
FT Modified-site 406 /note= "potential O-linked glycosylation site"
FT Modified-site 433 /note= "potential O-linked glycosylation site"
FT Modified-site 435 /note= "potential O-linked glycosylation site"
FT Modified-site 441 /note= "potential O-linked glycosylation site"
FT Modified-site 442 /note= "potential O-linked glycosylation site"
FT Modified-site 443 /note= "potential O-linked glycosylation site"
FT Modified-site 445 /note= "potential O-linked glycosylation site"
FT Modified-site 470 /note= "potential O-linked glycosylation site"
FT Modified-site 497 /note= "potential O-linked glycosylation site"
FT Modified-site 527 /note= "potential O-linked glycosylation site"
FT /note= "potential O-linked glycosylation site"

FT Modified-site 529 /note= "potential O-linked glycosylation site"
 FT Modified-site 537 /note= "potential O-linked glycosylation site"
 FT Modified-site 561 /note= "potential O-linked glycosylation site"
 FT Modified-site 565 /note= "potential O-linked glycosylation site"
 FT Domain 591..630 /note= "domain C"
 FT Modified-site 591 /note= "potential N-linked glycosylation site"
 FT Modified-site 592 /note= "potential N-linked glycosylation site"
 FT Modified-site 616 /note= "potential N-linked glycosylation site"
 FT Modified-site 630 /note= "potential N-linked glycosylation site"
 FT binding 688..731 /note= "potential N-linked glycosylation site; hyaluronan binding consensus sequence; part of domain D"
 FT Misc-difference 762 /note= "EGF-like domain"
 FT /note= "encoded by AAAA"
 PN WO200026367-A2.
 XX 11-MAY-2000.
 XX 29-OCT-1999; 99WO-US025440.
 XX 29-OCT-1998; 98US-00183972.
 PA (IOWA) UNIV IOWA RES FOUND.
 XX Hageman GS, Kuehn MH;
 XX WPI: 2000-365616/31.
 DR N-PSDB; AAA46205.
 XX Nucleic acids encoding interphotoreceptor matrix proteoglycans useful for preventing, diagnosing and treating ocular disorders such as retinal detachment and chorioretinal degeneration.
 PT Claim 5; Fig 3; 183pp; English.
 PS The present sequence represents an interphotoreceptor matrix (IPM) proteoglycan, designated IPM150. The protein is an IPM component (IPMC). Two subfamilies of IPMCs, IPM150 and IPM200, exist. The human IPM150 gene is located on chromosome 6q13-q15, between markers CHLC.GATALLF10 and D6S284. The IPM proteins may be used to supplement a patient's own production of the protein or to rectify alterations in their nucleic acids that result in expression of an inactive protein. The IPM nucleic acids may be used in this way to treat ocular diseases such as retinal detachment, chorioretinal degeneration, RPE (retinal pigment epithelium) degeneration, photoreceptor degeneration, age related macular degeneration, cone degeneration, mucopolysaccharidosis, rod-cone dystrophy and cone-rod dystrophy. The nucleic acids and proteins may also be used to assay for other modulators of IPM proteoglycan expression and activity that may be used to treat ocular diseases. The nucleic acids and proteins may also be used as diagnostic reagents to detect the presence of IPM nucleic acids and their products in samples from patients according to standard methodologies
 XX SQ Sequence 771 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 771
 Score: 3862.00 Matches: 769
 Percent Similarity: 96.74% Conservative: 2
 Best Local Similarity: 96.49% Mismatches: 0
 Query Match: 65.74% Indels: 26
 DB: 3 Gaps: 1

US-10-007-270-1 (1-3330) x AAY93336 (1-771)
 QY 131 ATGTATTGGAAACTAGAGAGCTATTTTGTGTTTGGATTTTCTCAAGTTCAGGA 190
 Db 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGluValGlnGly 20
 QY 191 ACCAAAGATATCTCATACATATACCATTCGAACTAAAGACATAGACATCCCCCA 250
 Db 21 ThrIleAspIleSerIleAsnIleTyrHisSerGluThrIleAspIleAspAsnPro 40
 QY 251 AGAATGAAACAACTGAAAGTACTGAAAAAATGTACAAAATCTCAACTATCAGCAATA 310
 Db 41 ArgAsnGluThrThrGluSerThrGluIleMetTyrIleMetSerThrMetArgIle 60
 QY 311 TTTCGATTGGCAAGCATCGAAACAAAGATCCGCATTTTCCCAAGGGGGTAAAGTC 370
 Db 61 PheAspLeuAlaIleValHisArgThrIleArgSerAlaPhePheProThrGlyValVal 80
 QY 371 TGTCCACAGGAATCCATGAAACAGATTTTACAGCTTTCAGCTTTCAGCTTATAGATTG 430
 Db 81 CysProGlnGluSerMetIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100
 QY 431 GTGTCTCAGGAAGCAGTATGGGAAGCATATCGGATCTTCTGGATCCCATCCCTGACACA 490
 Db 101 ValCysGlnGluAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120
 QY 491 GGGGAATATCAGGACTGGGTGAGCATCTGCCAGCAGGAGACCTTCTGCTCTTTGACATT 550
 Db 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140
 QY 551 GCAAAAACTTCCAGCAATTTCCAGGAGCACCCTGGATCTTCTCAGCAGAGATAAAGAC 610
 Db 141 GlyIleAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleIleGln 160
 QY 611 AGAAGTTTCCCTGACAGAAAAGATGAATATCTGCAGAGAGACATTTGGAGAGCCTGGT 670
 Db 161 ArgSerPheProAspArgIleAspGluIleSerAlaGluIleGluIleGluIleGluIle 180
 QY 671 GAAACCATGTCATTTTCAACAGCATCTACATTTCAAGACTTTGGCAGTATTTCAAGAA 730
 Db 181 GluThrIleValIleSer--Thr-----
 QY 731 AACCTCAGAGAGCAAAATTCAGATCTTGCACAGCTCTCAGTTGGGCTTTCCCTCTCA 790
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeu 200
 QY 791 CTCTGATGACACCTCTCTCAATGAATTCGATAATACACTCAACACACCAAGATCG 850
 Db 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrIleSer 220
 QY 851 CTAAACAGAGAGAAACAGATTCGCTGTGTGGAGGAGAGGCTGGAGCTCAGCG 910
 Db 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGluGluValGluLeuSerV 240
 QY 911 TCTCTCTGTAAACAGAGCTTCAAGGAGAGCTCGCTCAGCTCCAGCTCCCATATTACC 970
 Db 240 alSerLeuValAsnGlnIlePheIleAlaGluLeuAlaAspSerGlnSerProTyrTyG 260
 QY 971 AGAGCTAGCAGGAAGTCCCACTTCAGATGCAAAAGATATTAAAGAAATTTCCAGGAT 1030
 Db 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnIlePheIleGlyIleGlyIleProGly 280
 QY 1031 TCAAAAAATCCATGTGTTAGATTAGACCAAGAAAGAAAAAGATGGCTCAAGCTCCA 1090
 Db 280 helysIleHisValLeuGlyPheArgProLysIleGlyAspGlySerSerSert 300
 QY 1091 CAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTCGACAGCAAAAGCCCTGAA 1150
 Db 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAla 320
 QY 1151 GTGACCTCTCTGTTTGTATTCCAAACAAATTCAGAGTGGAGAGTCTATCATGGAACA 1210
 Db 320 erAspLeuLeuSerPheAspSerAsnIleGluSerGluGluValTyrHisGlyThrM 340

QY 1211 TGGAGGAGCAAGCAACAGAAATCTATCTACAGCTACAGACCTCAAAAGGCTGATCA 1270
 Db |||||
 QY 340 etGluGluAspLysGlnProGluLileThyLeuThrAlaThraspLeuLysArgLeuIleS 360
 |||||
 QY 1271 GCAAGCACTAGAGAGAGACAATCTTTGGATGTGGGACAATTCAGTTCTACTGATGAAA 1330
 Db erlySAlaLeuGluGluGlnSerLeuAspValGlyThyIleGlnPheThrAspGluI 380
 |||||
 QY 1331 TTGCTGGATCACTGCCAGCTTTGGTCTCTGACACCAATCAGAGCTGCCACATCTTTTG 1390
 Db leaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400
 |||||
 QY 1391 CTGTTTAAACAGAGATGCTACTTTGAGTCCAGAACTTCTCTCTGTTGAAACCCAGCTTG 1450
 Db laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420
 |||||
 QY 1451 AGACAGTGGAGGAGCAGACATGCTTACCTGACACTTTCTTGGTCTCCACCTGCTATGG 1510
 Db luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440
 |||||
 QY 1511 CCTTACCTCCCTCTCAGAGCTCCACCTTTCTTTATGGCATCAAGCATCTTCTCTGA 1570
 Db laSerThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT 460
 |||||
 QY 1571 CTGATCAGGCAACACAGATACAAATGGCCACTGACCAGACAATGCTAGTACAGGGCTCA 1630
 Db hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480
 |||||
 QY 1631 CCATCCCAACACAGTATTTCTGCAATCAGCCAACTGGCTCTGGAAATTCACATCCAC 1690
 Db hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500
 |||||
 QY 1691 CTGCATCTCAGATGACAGCCGATCAAGTGCAGGTCGGAGATATGTCAGACACTTAG 1750
 Db toAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520
 |||||
 QY 1751 ATGAAATGGATCTGTCTGCACTCTGCCCATCTGAGGTACAGAGCTCAGACCAATAG 1810
 Db spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540
 |||||
 QY 1811 TTTCTGCCACATCATTTCTGAGGATACCACTCTGTCTCAGCTTTACAGTATATCA 1870
 Db alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560
 |||||
 QY 1871 CCATAGTTCTATGACCATTTGCCCAAGCGGCGAGAGCTGTGACTGTCTTCAGTCTGC 1930
 Db hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580
 |||||
 QY 1931 GTGTCTCAACATGCCCTTCTCCAGCACTGTTCACAGAGCTCTCTGGAGTACCGAG 1990
 Db rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600
 |||||
 QY 1991 CTCTGGAGCAACAATTCACAGCTGCTGTCTCATATCTAGATCCATCTTACAGGAT 2050
 Db laLeuGluGlnGlnPheThrGlnLeuValProTyrLeuArgSerAsnLeuThrGlyP 620
 |||||
 QY 2051 TTAAGCACTTGAATATCTTAATCTCAGAAACGGAGTGTGATGTGATAGCAAAATCA 2110
 Db helyGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
 |||||
 QY 2111 AGTTTCTAAGTCTGCGGTATACCTTACCAAGCTGTGCAGGGGTCTTGGAGGATT 2170
 Db ysPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660
 |||||
 QY 2171 TTGTTTCTGTGAGCCCAACCACTCCATCTGGAATAGACAGCTACTCTCTCAACATTG 2230
 Db heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680
 |||||
 QY 2231 AACCACTGATCAGCAGATCCCTGCAAGTTCTTGGCTGGCGGAAATTTGCCAATGTG 2290
 Db luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
 |||||

QY 2291 TAAAGAACGAAACGAGTCTGAGAACCGAGTGTGCTGCAAAACAGGATATGACAGCCAGG 2350
 Db |||||
 QY 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720
 |||||
 QY 2351 GAGCCCTGGACGCTCTGGAACCGAGCTCTGTGGCCCTGGCACAAGAGSNAATGCCAGGTCC 2410
 Db lySerLeuAspGlyLeuGluProGlyLeuCysGly-LeuAlaGlnArgAsnAlaArgSer 739
 |||||
 QY 2411 TCCAGGGAAGGAGAGCTCCATGAGGTTGCCAGATCACTCTGAAATCAAGCATACAAA 2470
 Db SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyrIlyst 759
 |||||
 QY 2471 CTAGTGTAAAGTGTCCAAATCAACAAATAACAGG 2509
 Db hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771
 |||||

RESULT 5

ADA14867
 ID ADA14867 standard; protein; 771 AA.
 XX
 AC ADA14867;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human variant IPMC 150 isoform A.
 XX
 KW human; IPMC 150 isoform A; gene therapy;
 KW interphotoreceptor matrix component; IPMC; ocular disorder;
 KW macular degeneration; photoreceptor death; retinal detachment.
 XX
 OS Homo sapiens.
 XX
 PN US2002160954-AI.
 XX
 PD 31-OCT-2002.
 XX
 PF 08-NOV-2001; 2001US-00007270.
 XX
 PR 29-OCT-1998; 98US-00183972.
 PR 29-OCT-1995; 99US-00430195.
 XX
 PA (TOWA) UNIV IOWA RES FOUND.
 XX
 PI Hageman GS, Kuehn MH;
 DR WPI: 2003-238235/23.
 DR N-PSDB; ADA14866.
 XX
 PT New isolated or recombinant interphotoreceptor matrix component
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,
 PT treating or prognosticating ocular disorders, e.g. macular degeneration
 or retinal detachment.
 XX
 PS Claim 8; Page 66-68; 76pp; English.
 XX
 CC The invention relates to an isolated or recombinant interphotoreceptor
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)
 CC gene operatively linked to the IPMC polynucleotide. The IPMC
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,
 CC preventing, treating or prognosticating ocular disorders, e.g. macular
 CC degeneration, photoreceptor death or retinal detachment. They are also
 CC useful for identifying a compound capable of modulating IPMC gene
 CC expression in a cell. The present sequence represents the amino acid
 CC sequence of the variant human interphotoreceptor matrix component, IPMC,
 CC 150 isoform A.
 XX
 SQ Sequence 771 AA;

Alignment Scores:

Pred. No.: 0 Length: 771
 Score: 3862.00 Matches: 769
 Percent Similarity: 96.74% Conservative: 2

| Best Local Similarity: | 96.49% | Mismatches: | 0 |
|---|--------|--|------|
| Query Match: | 65.74% | Indels: | 26 |
| DB: | 6 | Gaps: | 1 |
| US-10-007-270-1 (1-3330) x ADA14867 (1-771) | | | |
| QY | 131 | ARGTATTTGGAACTAGAGAGCTATTTTGTGTTTTCCTCAAGTTCAAGGA | 190 |
| Db | 1 | MetYrLeuGluThrArgArgAlaIlePheValPheThrIlePheLeuGlnValGlnGly | 20 |
| QY | 191 | ACCAAGATATCTCCATTAACTATACCATCTGAACTAAAGACATAGACAATCCCCCA | 250 |
| Db | 21 | ThrIysAspIleSerIleAsnIleTyHisSerGluThrIysAspIleAspAsnProPro | 40 |
| QY | 251 | AGAAATCAAACTCAAGTACTGAAAATGTACAAATGTCAAAATGTCACTATGAGACGA | 310 |
| Db | 41 | ArgAsnGluThrThrGluSerThrGluYsMetYrIysMetSerThrMetArgArgIle | 60 |
| QY | 311 | TTGCAATTTGGCAAGCATTCGAACAAAGATCCGCACTTTTCCCAACGGGGTTAAAGTC | 370 |
| Db | 61 | PheAspLeuAlaIysHisArgThrIysArgSerAlaPheProThrGlyValIysVal | 80 |
| QY | 371 | TGTCACAGCAATCCATCAACACAGATTTAGACAGTCTTCAAGCTTATTATAGATTGAG | 430 |
| Db | 81 | CysProGlnGluSerMetIysGlnIleLeuAspSerLeuGlnAlaTyTyArgLeuArg | 100 |
| QY | 431 | GTGTGTCAGAGCAGTATGGGAAGCATATCGCATCTGCGACGAGACCTTCTGCCTCTT | 490 |
| Db | 101 | ValCysGlnGluAlaValTrpGluAlaTyArgIlePheLeuAspArgIleProAspThr | 120 |
| QY | 491 | GGGGAATATCAGGACTGGTCTGACATCTGCGACGAGACCTTCTGCCTCTTTCACATT | 550 |
| Db | 121 | GlyGluTyGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle | 140 |
| QY | 551 | GGAAAAAATCTTCACAAATCCACAGACACTCGATCTTCTCCAGCAGAGAATAAAACAG | 610 |
| Db | 141 | GlyIysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleIysGln | 160 |
| QY | 611 | AGAGTTTCCCTGACAGAAAGATCAATATCTGCAGAGAGACATTCGGACAGCCTGGT | 670 |
| Db | 161 | ArgSerPheProAspArgIysAspGluIleSerAlaGluIysThrLeuGlyGlnProGly | 180 |
| QY | 671 | GAACCACTTGCTATTCACAGCAATCTCATTTCAAAGACTTGGCGAGTATTTCAAGAA | 730 |
| Db | 181 | GluThrIleValIleSer--Thr----- | 187 |
| QY | 731 | AACTCTCAGAGAGCAAAATCAAGATCTGCCAAGCTCTACCTGGGCTTTCCCTCTCA | 790 |
| Db | 188 | -----AspValAlaAsnValSerLeuGlyProPheProLeuT | 200 |
| QY | 791 | CTCTCGATGACACCTCTCAATGAAATCTCGATAATACACTCAACACACACCAAGATGC | 850 |
| Db | 200 | hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrIysMetP | 220 |
| QY | 851 | CTACAAACAGAAAGAAACAGAAATTCGCTGTGTGGAGGACGAGGGTGGAGCTCAGCG | 910 |
| Db | 220 | roThrThrGluArgGluThrGluPheAlaValLeuGluGlnArgValGluLeuSerV | 240 |
| QY | 911 | TCTCTCTGGTAAACAGAAAGTTCAAGCGACAGCTCGCTGACTCCCACTGCCCATTTACC | 970 |
| Db | 240 | alSerLeuValAsnGlnIysPheLysAlaGluLeuAlaAspSerGlnSerProTyTyG | 260 |
| QY | 971 | AGGAGCTACAGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAAATTTCCAGAT | 1030 |
| Db | 260 | lnGluLeuAlaGlyIysSerGlnLeuGlnMetGlnIysIlePheIysIysLeuProGlyP | 280 |
| QY | 1031 | TCAAAAAATCCATGTGTAGGATTTAGACCAAGAAAGAAAAAGATGGCTCAAGTCCA | 1090 |
| Db | 280 | helyIysIleIleHisValLeuGlyPheArgProIysIysGluIysAspGlySerSerSert | 300 |
| QY | 1091 | CAGATGCAACTTACGGCCATCTTTAAGAGACACAGTGCAGACGCAAAAAGCCCTGCA | 1150 |
| Db | 300 | hrGluMetGlnLeuThrAlaIlePheIysArgHisSerAlaGluAlaIysSerProAlaS | 320 |

QY 2231 AACGAGCTGATCAGACGATCCCTGCAAGTTCCTGCGCTGGCGGCGAATTTGCCCAATGTG 2290
 Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
 QY 2291 TAAAGAACGAACGAGTCTGAGGAGCGGAGTGTGCTGCAACACCGAGGATATGACACCGAGG 2350
 Db 700 allyAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyAspSerGlnG 720
 QY 2351 GGAGCTCGAGCGGTCTGGAAACAGGCTCTGTGGCCCTGGCACAAAGGATGCGAGGTCC 2410
 Db 720 lysLeuAspGlyLeuGluProGlyLeuGlyLeuAlaGlnArgAsnAlaArgSer 739
 QY 2411 TCCAGGAAAGAGGCTCCATCCAGCTGTCAGATCACTCTGAAATCAAGCATCAAAA 2470
 Db 740 SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyLysT 759
 QY 2471 CTAGTGTAAAAAGTTCCTCAAAATCAACAAAATAACAAGG 2509
 Db 759 hrSerVal-LysSerLysLysLeuAlaThrArg 771
 RESULT 6
 ID ABR42354 standard; protein; 771 AA.
 XX AC ABR42354;
 XX DT 11-AUG-2003 (first entry)
 XX DE Human interphotoreceptor matrix IPM 150, isoform A variant.
 XX KW Human; interphotoreceptor matrix; IPM 150, IPMC; receptor;
 XX KW ophthalmological; gene therapy.
 XX OS Homo sapiens.
 XX FN WO2003039346-A2.
 XX PD 15-MAY-2003.
 XX PF 08-NOV-2002; 2002WO-US036090.
 XX PR 08-NOV-2001; 2001US-00077270.
 XX PA (IOWA) UNIV IOWA RES FOUND.
 XX PI Hageman GS, Kuehn MH;
 XX DR MPI; 2003-441440/41.
 XX DR N-PSDB; ACC57960.
 XX PT New interphotoreceptor matrix proteins and polynucleotides, useful for
 PT treating or preventing photoreceptor death or retinal detachment, or for
 PT treating ocular disorders.
 XX PS Claim 6; Page 94-96; 105pp; English.
 XX CC The present sequence is the protein sequence of a variant of isoform A of
 CC novel human interphotoreceptor matrix 150 (IPM 150), a member of the
 CC newly identified interphotoreceptor matrix component (IPMC) gene family.
 CC The variant lacks the last 25 amino acids of the IPM 150 isoform A.
 CC that are not present in other sequence. he has 9 residues at the C-terminus
 CC that are not present in other sequence. he IMP 150 gene is located on
 CC chromosome 6q13-q15, a region that also contains loci for progressive
 CC bifocal chorioretinal atrophy, autosomal dominant Stargardt's-like
 CC macular dystrophy, North Carolina macular dystrophy and Salla disease.
 CC Members of the IPMC gene family have been identified in humans, monkey,
 CC cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2 Subfamilies are
 CC designated IPM 150 (or IPMG1) and IPM 200 (or IPMG2). The invention
 CC provides IPM 150 and IPM 200 polynucleotides and polypeptides, antibodies
 CC that specifically bind the polypeptides, and vectors comprising the
 CC polynucleotides. A claimed method of treating or preventing photoreceptor
 CC death or retinal detachment involves administering an IPMC
 CC polynucleotide, polypeptide or antibody. Also claimed is a method for

CC identifying a compound capable of modulating IPMC gene expression
 XX Sequence 771 AA;
 SQ
 Alignment Scores:
 Pred. No.: 0 Length: 771
 Score: 3862.00 Matches: 769
 Percent Similarity: 96.74% Conservat: 2
 Best Local Similarity: 96.49% Mismatches: 0
 Query Match: 65.74% Indels: 26
 DB: 7 Gaps: 1
 US-10-007-270-1 (1-3330) x ABR42354 (1-771)
 QY 131 ATCTATTTGGAACTCAGACGATATTTTGTGTTTGGATTTTCTCCAGTTCAGGA 190
 Db 1 MetTyLeuGluThrArgArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCTCCATTAAACATATACCATTTCTGAAACTTAAAGACATAGACAATCCCCA 250
 Db 21 ThrLysAspIleSerIleAsnIleTyHisSerGluThrLysAspIleAspAsnProPro 40
 QY 251 AGAATGAACAACACTGAAAGTACTGAAAAATGTACAAAATGTCAATATGAGACGAATA 310
 Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyLysMetSerThrMetArgArgile 60
 QY 311 TTCGATTTGGCAAGCATCGAACAAAAGATCCGCAATTTTCCCAACGGGGTTAAAGTC 370
 Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
 QY 371 TGTCCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATATATAGATGAGA 430
 Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyTyArgLeuArg 100
 QY 431 GTGTGTGAGGACAGCATATGGAGAGCATATCGGATCTTCTGGATCCCATCCCTGACACA 490
 Db 101 ValCysGlnGluAlaValTrpGluAlaTyArgIlePheLeuAspArgileProAspThr 120
 QY 491 GGGGAATATCAGGACTGGTCTGAGCATCTGCAGCAGGAGACCTTCTGCTCTTTCACATT 550
 Db 121 GlyLysGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140
 QY 551 GGAATAAATCTCAGCAATTTCCAGGACACTTGGATCTTCTCCAGCAGAGATAAAGACAG 610
 Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 160
 QY 611 AGAAGTTTCCCTCAGCAAGAAAGATGAATATCTGCAGAGAGAGACATTGGGAGAGCTGGT 670
 Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluProGly 180
 QY 671 GAAACCATTTGCTATTTCAACAGCAATCTACATTTCAAAGACTTGGGAGCTATTCTAAGAA 730
 Db 181 GluThrIleValIleSer--Thr----- 187
 QY 731 AACCTCAGAAGAGCAAAATTCAGATGTTGCCAAGCTCTCACTTGGGCGCTTTCCCTCTCA 790
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeuT 200
 QY 791 CTCCTGTATGACACCTCTCATTCAATCAATCTTCTGATATATACACTCAACAGACACCAAGATGC 850
 Db 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 220
 QY 851 CTACACAGAAAGAGAAACAGAAATTCGCTGTGTGGAGGAGAGAGGGTGGAGCTCAGCG 910
 Db 220 toThrThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 240
 QY 911 TCTCTCTGTGTAAACAGAGAGTTCAGGCGAGAGCTCGTGTGACTCCAGTCCCCCATATTACC 970
 Db 240 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyTyArg 260
 QY 971 AGGAGCTAGCAGGAAAGTCCCACTTCAGATTCAAAAGATATTATTAAGAACTTCAGAT 1030
 Db 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGly 280

QY 1031 TCAGAAATCCATGCTGTAGCATTTAGACCAAGAGAAAGAAAGATGCTCAAGCTCCA 1090
Db |||||
QY 280 heLysLysIleHisValLeuGlyPheArgProLysLysGlyLysAspGlySerSerSert 300
Db |||||
QY 1091 CAGAGATGCAACTTACGGCCCATCTTTAAGAGACACAGTCAGAGCAAAAGACCCCTCAAA 1150
Db |||||
QY 300 hrGluwetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlaS 320
QY 1151 GTGACCTCCCTGCTTTGATTCCCAAAATTTGAAAGTGGAGAGTCTATCATCGAACCA 1210
Db 320 arAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 340
QY 1211 TGGAGGAGGACAGCAACCAAGATCTATCTCAGAGTACAGACTCAGAGGCTGATCA 1270
Db 340 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 360
QY 1271 GCAAGCACTAGAGGAGCAACAATCTTTGGATGTGGGACAAATTCAGTTCACTGATGAAA 1330
Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380
QY 1331 TTGCTGATCAGCTGCCAGCTTTGCTCTGACACCCCAATCAGAGCTGCCACATCTTTTG 1390
Db 380 leaLagLysLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400
QY 1391 CTGTTATACAGAGGATGCTACTTTGAGTCCAGAACTTCTCTGTTGACCCAGCTTG 1450
Db 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420
QY 1451 AGACAGTGGACGGACGACGATGCTTACCTGACACTTCTGCTCTCCACCTGCTATGG 1510
Db 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440
QY 1511 CCTCTACCTCCCTCTGAGAGCTCCACTTTCTTTATGCGATCAAGCATCTTCTCTGA 1570
Db 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeuT 460
QY 1571 CTGATCAAGGACACACAGATACAAATGGCCACTGACACAGACATGCTAGTACAGGCTCA 1630
Db 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480
QY 1631 CCATCCCAACAGTGTATTCTGCAATTCAGCACTGCTCTGGGAATTCATCATCAC 1690
Db 480 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaGlyLysSerHisProp 500
QY 1691 CTGCACTTTCAGATGACAGCCATCAGTCCAGTGGCGAGATATGCTCAGACACCTAG 1750
Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 520
QY 1751 ATGAAATGGATCTGTCTGACACTCTGCCCCATCTGAGTACACGAGCTCAGCGAATAG 1810
Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540
QY 1811 TTTCTGTCCAGATCATTTCTGGAGGATACCACTCTGCTCAGCTTTACAGTATATCA 1870
Db 540 aLSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560
QY 1871 CCACTAGTTCATGACCATTCGCCCAAGGCGGAGACTGCTAGTGTCTTCAGTCTGC 1930
Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 580
QY 1931 GTGTGTCAACAGGCTCTCTCCAAAGCACTGTCTCAACAGAGCTCTCTGGGTACCGAG 1990
Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600
QY 1991 CTCTGGAGCAACATTCACAGCTGCTGTTCTCCATATCTACGATCCATCTTACAGAT 2050
Db 600 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620
QY 2051 TTAAGCAACTTGAATCTTAATCTTCAAGAAAGGAGTGTGATGTGAATAGCAAAATGA 2110
Db 620 heLysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640

QY 2111 AGTTTGTAAGTCTGTGCGGTATACCTTCAACAGGCTGTGACGCGGTCTTCGAGGATT 2170
Db 640 yPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660
QY 2171 TTCGTTCTGCTGACGCCCAACAACTCCATCTTGAAATAGACAGCTACTCTCTCAATTG 2230
Db 660 heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680
QY 2231 ACCAGCTGTATCAAGCAGATCCCTGCAAGTTCCTGCGTGGCGCAATTTGCCCAATGTG 2290
Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
QY 2291 TAAAGACCAACGCACTGAGGAGCGGAGTCTGCTGCAAAACCAAGATATGACCCAGG 2350
Db 700 alLysAsnGlnArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 720
QY 2351 GGAGCTGGAGCGTCTTGGAAACAGGCTCTGTGGCCCTGTGGCCACAAAGGAATCCGAGGTCC 2410
Db 720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyLeuAlaGlnArgAsnAlaArgSer 739
QY 2411 TCCAGGAAAGGAGGCTCCATGCGAGTTCGAGTTCGCAAAATCAAGCATACAAAA 2470
Db 740 SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyrLysT 759
QY 2471 CTAGTGTAAAGTTCCAAAATCAAAAATTAACAAGG 2509
Db 759 hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771
RESULT 7
ADA14843
ID ADA14843 standard; protein; 719 AA.
XX
AC ADA14843;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human interphotoreceptor matrix component, IPMC, 150 isoform B.
XX
KW human; IPMC 150 isoform B; gene therapy;
KW interphotoreceptor matrix component; IPMC; ocular disorder;
KW macular degeneration; photoreceptor death; retinal detachment.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..20 /note= "Signal sequence"
FT Region 1..7 /note= "Not encoded by {seqid:3}"
FT Protein 21..719 /note= "Mature IPMC 150 isoform B"
XX
PN US2002160954-A1.
XX
PD 31-OCT-2002.
XX
PF 08-NOV-2001; 2001US-00007270.
XX
PR 29-OCT-1998; 98US-00183972.
XX
PR 29-OCT-1999; 99US-00430195.
XX
PA (IOWA) UNIV IOWA RES FOUND.
XX
PI Hageman GS, Kuehn MH;
XX
DR WPI; 2003-238235/23.
XX
DR N-PSDB; ADA14842.
XX
PT New isolated or recombinant interphotoreceptor matrix component
PT polynucleotide and polypeptide, useful for diagnosing, preventing,
PT treating or prognosticating ocular disorders, e.g. macular degeneration
PT or retinal detachment.
XX

PS Claim 8; Page 34-36; 76pp; English.

XX The invention relates to an isolated or recombinant interphotoreceptor
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)
 CC gene operatively linked to the IPMC polynucleotide. The IPMC
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,
 CC preventing, treating or prognosticating ocular disorders, e.g. macular
 CC degeneration, photoreceptor death or retinal detachment. They are also
 CC useful for identifying a compound capable of modulating IPMC gene
 CC expression in a cell. The present sequence represents the amino acid
 CC sequence of human interphotoreceptor matrix component, IPMC, 150 isoform
 XX B.

SQ Sequence 719 AA;

Alignment Scores:

Pred. No.: 0 Length: 719
 Score: 3611.00 Matches: 717
 Percent Similarity: 87.67% Conservative: 1
 Best Local Similarity: 87.55% Mismatches: 1
 Query Match: 61.46% Indels: 100
 DB: 6 Gaps: 2

US-10-007-270-1 (1-3330) x ADA14843 (1-719)

QY 131 ATGTATTTGGAACTAGAGAGCTATTTTGGTATTTTGGATTTTCCAAAGTCAAGGA 190
 DB 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCTCCATTACATATACCATTTCTGAACTTAAGACATAGACATCCCCCA 250
 DB 21 ThrLys----- 22
 QY 251 AGAATGAACAACGTGAAGTACTGAAAAATGTACAAAATGTCAACTATGAGACGAATA 310
 DB 22 ----- 22
 QY 311 TTGGATTTGGCAAGCATCGAACAAAAGATCCGATTTTCCACGGGGTTAAAGTC 370
 DB 22 ----- 22
 QY 371 TGTCACAGGAATCCATAAGACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430
 DB 22 ----- 22
 QY 431 GTGTCTCAGGAGCAGTATGGAGCATATCGGATCTTCTGGATCGCATCCCTGACACA 490
 DB 23 ValCysGlnGluAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 42
 QY 491 GGGGAATATCAGGACTGGCTCAGCATCTGCCAGCAGGAGACCTTCTGCTCTTGCATT 550
 DB 43 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 62
 QY 551 GGAATAATCTCAGCAATTTCCAGGAGCAGCTGGATCTTCTCCAGCAGAGATAAAGCAG 610
 DB 63 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnArgIleLeuGln 82
 QY 611 AGAAGTTTCCCTCAGCAAGAGATGAATATCTGCAGAGACATTTGGGAGAGCTGGT 670
 DB 83 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluProGly 102
 QY 671 GAAACCATTTGCATTTTCAACAGCAATCTACATTTTCAAGACTTGGGAGTATCTAAGAA 730
 DB 103 GluThrIleValIleSer--Thr----- 109
 QY 731 AACCTCAGAGAGCAAAATTCAGATGTTGCCAAGCTTCTCATTGGGCTTTCCCTCTCA 790
 DB 110 -----AspValAlaAsnValSerLeuGlyProPheProLeu 122
 QY 791 CTCCTGATGACACCTCTCCTCAATGAATTTCTCGATATATACACTCAACGACACCAAGATGC 850
 DB 122 hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP 142

QY 851 CTACACACAGAAAGAGAAACAGAAATTCGCTGTGTTGGAGGAGCAGAGGTTGAGCTCAGCG 910
 DB 142 rofhrThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 162
 QY 911 TCTCTCTGGTAAACACAGAGTTCAAGGCGAGAGCTCGCTGACTCCAGTCCCATATTACC 970
 DB 162 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTrpG 182
 QY 971 AGGAGCTACAGGAAGTCCCAACTTCAGATGCAAAAGATATTTAAGAACTTCCAGAT 1030
 DB 182 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLeuProGlyP 202
 QY 1031 TCAAAAAAATCCATGTTAGGATTTAGACCAAAAGAAAGAAAGATGCTCAAGCTCCA 1090
 DB 202 heLysLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerS 222
 QY 1091 CAGAGATGCACTTACGGCCATCTTTAAGACACACAGTCGAGAACAAAAAGCCCTGCAA 1150
 DB 222 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAla 242
 QY 1151 GTGACCTCTGCTCTTTGATTTCCAAAAATTTGAAAGTGAAGTGAAGTCTTATCATGGAACA 1210
 DB 242 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 262
 QY 1211 TGGAGGAGCAGACCAACCAAGAAATCTATCTCAGACTCAGACCTCAAAAGGCTGATCA 1270
 DB 262 etGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuL 282
 QY 1271 GCAAGCACTAGAGAGAAACAATCTTTGGATGTGGGGAACAATTCAGTTCACATGATGAAA 1330
 DB 282 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGlu 302
 QY 1331 TTCTGTGATCACTCCAGGCTTTGCTCTGACACCAATCAGAGTCCACACATCTTTG 1390
 DB 302 leaLysSerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 322
 QY 1391 CTGTTAATACAGAGATGCTTACTTTGAGTCCAGAACTTCTCTCTTGAACCCAGCTTG 1450
 DB 322 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 342
 QY 1451 AGCAGTGGACGGACGAGCATGTCTTACTGACACTTCTTGGTCTCCACCTGCTATGG 1510
 DB 342 luthrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMeta 362
 QY 1511 CCTCTACCTCTCTCAGAGGCTCCACCTTTCTTATGTCATCAGCATCTTCTCTCTGA 1570
 DB 362 laSerThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeu 382
 QY 1571 CTGATCAAGGCAACACAGATACAATGGGCACTGACCAAGATGCTAGTACCAAGGCTCA 1630
 DB 382 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeu 402
 QY 1631 CCATCCCAACAGTGAATTCCTGCAATCAGCACTGCTGGCTCTGGAAATTCACATCCAC 1690
 DB 402 hrIleProThrSerAspTrpSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 422
 QY 1691 CTGCACTTTCAGATGACAGCCCATCAAGTGGGAGATATGCTCAGACACCTAG 1750
 DB 422 roAlaSerSerAspAspSerArgSerSerAlaGlyGluAspMetValArgHisLeuA 442
 QY 1751 ATGAAATGATCTGTCTGCACTCTCTGCCCATCTGAGGTACAGAGCTCAGCAATATG 1810
 DB 442 spGluMetAspLeuSerAspThrProAlaProSerGluValProGlyLeuSerGluTyrV 462
 QY 1811 TTCTGTCCAGCATCTTCTTGGAGGATACCACTCTCTCTCAGCTTACAGTATATCA 1870
 DB 462 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIle 482
 QY 1871 CCATGATTTCTATGACCATTCGCCCAAGGCGCAGAGCTGGTAGTGTCTTCTCAGTCTGC 1930
 DB 482 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValPhePheSerLeuA 502

QY 1931 GTCTGCTAAATGCTGCTCTCTGAGAGCTGTTTCAACAGAGCTCTCTGAGTACCGAG 1990
 Db 502 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyArgA 522
 QY 1991 CTCTGAGCAACAATTCACACAGCTGCTGTTTCCATATCTACGATCCCAATCTTACAGAT 2050
 Db 522 laLeuGluGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 542
 QY 2051 TTAAGCACTTGAATTAATTAATCTTCAAGAACGGAGTGTGATTTGTGAATAGCAAAATGA 2110
 Db 542 helyeGlnLeuGluLeuLeuAsnPheArgaAnglySerValileValAsnSerLysMetL 562
 QY 2111 AGTTTCTTAAGTCTGCGCTATACCTTCAACAGCTGTGCAAGGGTCTTGAGGAT 2170
 Db 562 yPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 582
 QY 2171 TTCGTTCTGTCGAGCCCAACATCTCCATCTGGAATAGACAGCTACTCTCAACATTG 2230
 Db 582 heArgSerAlaAlaGlnGlnLeuHisLeuGluLeuAspSerTyrSerLeuAsnIleG 602
 QY 2231 AACGAGTATCAAGAGATCTCTGCAAGTTCCTGCGCTGCGCGAATTTGCCAATGTG 2290
 Db 602 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 622
 QY 2291 TAAAGAACGAGCTGAGGAGCGAGTGTGCTGCTGCAACAGAGTCTGCAAGGATCGAGTCC 2350
 Db 622 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 642
 QY 2351 GGAGCTGAGCGTCTGGAACAGAGCTCTGCTGCGCTGCGCAACAGGATCGAGTCC 2410
 Db 642 lySerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluValL 662
 QY 2411 TCAGGGAAGGAGCTCCATGCGAGTGTGCGAGTCACTTGAATAATCAAGCATACAAA 2470
 Db 662 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLysT 682
 QY 2471 CTAGTGTAAAAAGTTCCAAATCAACAAATAACAAGTAATCACTTAAAGAAATCTG 2530
 Db 682 hrSerValLysLysPheGlnAsnGlnGlnAsnAsnLysValileSerLysArgAensG 702
 QY 2531 AATTACTGCGCTAGAAATATGAAGATTTAAACCATCAAGATTGGAAGCAAAAT 2583
 Db 702 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 719

RESULT 8
 ABR42343
 ID ABR42343 standard; protein; 719 AA.
 XX AC ABR42343;
 XX DT 11-AUG-2003 (first entry)
 XX DE Human interphotoreceptor matrix IPM 150, isoform B.
 XX KW Human; interphotoreceptor matrix; IPM 150; IPMG; receptor;
 XX KW ophthalmological; gene therapy.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Signal_peptide
 FT Protein 21..719
 FT /label= IPM 150
 FT Domain 497..568
 FT /note= "conserved domain"
 FT Domain 586..628
 FT /note= "EGF-like domain"
 PN HQ2003039346-A2.
 XX PD 15-MAY-2003.
 XX

PF 08-NOV-2002; 2002WO-US036090.
 XX 08-NOV-2001; 2001US-00077270.
 XX (IOWA) UNIV IOWA RES FOUND.
 XX Hageman GS, Kuehn MH;
 XX WPI; 2003-441440/41.
 XX
 XX New interphotoreceptor matrix proteins and polynucleotides, useful for
 XX treating or preventing photoreceptor death or retinal detachment, or for
 XX treating ocular disorders.
 XX Claim 6; Page 79; 105pp; English.
 XX
 XX The present sequence is the protein sequence of isoform B of novel human
 XX interphotoreceptor matrix 150 (IPM 150), a member of the newly identified
 XX interphotoreceptor matrix component (IPMC) gene family. The IPM 150 gene
 XX is located on chromosome 6q13-q15, a region that also contains loci for
 XX progressive bifocal choriorretinal atrophy, autosomal dominant Stargardt's
 XX -like macular dystrophy, North Carolina macular dystrophy and Salla
 XX disease. Members of the IPMC gene family have been identified in humans,
 XX monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2
 XX Subfamilies are designated IPM 150 (or IPMG1) and IPM 200 (or IPMG2). The
 XX invention provides IPM 150 and IPM 200 polynucleotides and polypeptides,
 XX antibodies that specifically bind the polypeptides, and vectors
 XX comprising the polynucleotides. A claimed method of treating or
 XX preventing photoreceptor death or retinal detachment involves
 XX administering an IPMC polynucleotide, polypeptide or antibody. Also
 XX claimed is a method for identifying a compound capable of modulating IPMC
 XX gene expression
 XX SQ Sequence 719 AA;
 XX
 XX Alignment Scores:
 XX Pred. No.: 0 Length: 719
 XX Score: 3611.00 Matches: 717
 XX Percent Similarity: 87.67% Conservative: 1
 XX Best Local Similarity: 87.55% Mismatches: 1
 XX Query Match: 61.46% Indels: 100
 XX DB: 7 Gaps: 2
 XX
 XX US-10-007-270-1 (1-3330) x ABR42343 (1-719)
 QY 131 ATGTATTTGGAAGTAAAGAGTATTTTGTGTTTTCGATTTTCTCCAGTTCAAGA 190
 Db 1 MetTyrLeuGluThrArgArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAAGATATCTCCATTAAATATATACCATTTCTGAAACTAAAGACATAGACAAATCCCCA 250
 Db 21 ThrLys----- 22
 QY 251 AGAAATGAAACAACACTGAAAGTACTGAAATAATGTACAAATGTCACTATGAGAGCAATA 310
 Db 22 ----- 22
 QY 311 TTGATTTGGCAAGCATCGAACAAAGATCGCATTTTTCCTCCAGGGGGTTAAAGTC 370
 Db 22 ----- 22
 QY 371 TGTCCACAGGAATCCATGAAACAGATTTTATAGACAGTCTTCAAGCTTATTATAGATTGAGA 430
 Db 22 ----- 22
 QY 431 GTGTGTGCAAGACAGTATGGAGAGCATATCGGATCTTTCTGGATCGATCCCTGACACA 490
 Db 23 ValCysGlnGluAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 42
 QY 491 GGGGAATATCAGGACTGGGTCTGAGCATCTCCAGCAGGAGACCTTCTGCTCTTTCGACATT 550
 Db 43 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 62

QY 551 GAAAAAATTCAGCAATTCAGGAGCACCTGGATCTTCTCCAGCAGAGAAATAAACAG 610
 Db |||||
 QY 611 AGAAGTTTCCTGCAGAGAAAGATGAATATCTGCAGAGAGACATTTGGGAGACCTGGT 670
 Db |||||
 QY 671 GAACCAATTTGCTATTTCAACAGCAATCTACATTTCAAGAGCTTGGGAGTATCTTAAGAA 730
 Db |||||
 QY 731 AACCTCAGAGAGCAAAATTCAGATGTTGGCAAGTCTCACTTGGGCTTTCCTCTCA 790
 Db |||||
 QY 791 CTCTGATGACACCTCTCAATGAATTTCTGATATATACATCACTCAACGACCAAGATGC 850
 Db |||||
 QY 851 CTACACAGAGAGAAACAGAAATTCGCTGTTGGAGGAGCAGAGGCTGAGCTCAGCG 910
 Db |||||
 QY 911 TCTCTCTGTAACCCAGAGATTCAGGAGAGCTCGCTGACTCCAGTCCCATATATACC 970
 Db |||||
 QY 971 AGAGCTAGCAGAGAAAGTCCCAACTTCAGATGCAGAAAGATTTTAAAGATTCAGGAT 1030
 Db |||||
 QY 1031 TCAAAAAATCCATGTTTGGATTTAGACCAAGAGAAAGAAAGATGCTCAAGCTCCA 1090
 Db |||||
 QY 1091 CAGAGATGCACTTACGGCCATCTTTAAGAGACAAGTGCAGAGCAAGCAAAAGCCCTCAA 1150
 Db |||||
 QY 1151 GTGACCTCTCTGTTTGTATTCACCAAAATTTGAAGTGGAGAGTCTATCATGGAACCA 1210
 Db |||||
 QY 1211 TGGAGGAGGACAGCAACAGAAATCTATCTCAGAGTACAGACCTCAAAAGGCTGATCA 1270
 Db |||||
 QY 1271 GCAAGCACTAGAGAGCAACATCTTGGATGGGAGCAATTCAGTTCACATGATGAAA 1330
 Db |||||
 QY 1331 TTGCTGGATTCAGTCCAGCCTTTGGTCTGACACCCCAATCAGAGCTGCCACATCTTTG 1390
 Db |||||
 QY 1391 CTGTATACAGAGATGCTATTTGAGTCCAGAACTTCTCTCTGCAACCCAGCTTG 1450
 Db |||||
 QY 1451 AGACGTGAGGAGAGCAGCATGCTTACCTGACATCTTGTGCTCCACCTGCTATGG 1510
 Db |||||
 QY 1511 CTTCTACCTCTCTCAGAGCTCCACTTCTTATGTCATCAGCATCTTCTCTGA 1570
 Db |||||
 QY 1571 CTGATCAGGACACAGATACATGAGCCACTGACAGCAATGCTAGTACCAAGGCTCA 1630
 Db |||||
 QY 1631 CCATCCCCCAGTGTATTCTGCAATCAGCCACCTGGCTCTGGGAATTTTCATCCAC 1690

Db |||||ProThrSerAspLysSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisPro 422
 QY 1691 CTGCATCTTCAGATCAGACCGCATCAAGTGCAGGTGGCGAAGATATGTCAGACACCTAG 1750
 Db |||||SerSerAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA 442
 QY 1751 ATGAAATGGATCTGTCAGACATCTCCCTGCCCATCTGAGGTACCCAGAGCTCAGCGAATATG 1810
 Db |||||MetAspLeuSerAspThrProAlaProSerGluValProGlyLeuSerGluTyrV 462
 QY 1811 TTTCTGTCCCATGATCATTTCTTGGAGATACCACTCTCTGCTCTCAGCTTTCAGATATCA 1870
 Db |||||ValProAspHisPheLeuGluAspThrProValSerAlaLeuGlnTyrIleF 482
 QY 1871 CCCTAGTCTATGATCACTATGCCCCCAAGGCCGAGAGCTGTGTAGTCTTCTCAGTCTGC 1930
 Db |||||ThrSerSerMetThrIleAlaProLysGlyArgGlnLeuValValPhePheSerLeuA 502
 QY 1931 GTGTCTTAACATGCTCTCTCCACGACCTGTTCAACAGAGCTCTCTGGAGTACCAG 1990
 Db |||||ValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 522
 QY 1991 CTCTGAGCAACAATTCACACAGCTGCTGTGTCCATATCTACGATCCAAATCTTACAGAT 2050
 Db |||||LeuGlnGlnPheThrGlnLeuLeuValProLysLeuArgSerAsnLeuThrGlyP 542
 QY 2051 TTAAGCACTGAATACTTAATCTTCAGAAACGGAGTGTGATGTGAATAGCAAAATGA 2110
 Db |||||GlnLeuGluLeuLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 562
 QY 2111 AGTTGTCTAAGTCTGTGCGCTATTAACCTCACAGGCTGTGCACGGGTCTTGGAGGAT 2170
 Db |||||PheAlaLysSerValProLysLeuLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 582
 QY 2171 TTCGTTCTGTGAGCCCAACCACTCTGGAATAGACAGCTACTCTCTCAACATTG 2230
 Db |||||SerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 602
 QY 2231 AACCACTGATCAGCAGATCTCTGCAAGTCTGCGCTGCGCGAATTTGCCCAATGTG 2290
 Db |||||ProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 622
 QY 2291 TAAAGAACGAAACGACTCAGAGAGCGAGTGTGCTGCAACAGGATATGACAGCCAGG 2350
 Db |||||LysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 642
 QY 2351 GGACCTGAGACGCTGTGGAACAGGCTCTGTGCGCTGCGCAAGAGGAATGCGAGGTCC 2410
 Db |||||LysLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluValL 662
 QY 2411 TCAGGGAAGGAGCTCCATGCAAGTGTGCCAGATCTACTCTGAAAATCAAGCATACAAAA 2470
 Db |||||GlnGlyLysGlyAlaProCysArgLeuProAspHisSerGlnAsnGlnAlaTyrLysT 682
 QY 2471 CTAGTGTAAAAGTTCAAAATCAACAAAATAAGGTAATCAGTAAAGAAATCTG 2530
 Db |||||SerValLysLysPheGlnAsnGlnAsnAsnLysValIleSerLysArgAsnSerG 702
 QY 2531 AATTACTGACCTGAGATATGAGAAATTTAACCATCAAGATTTGGAGGAAAT 2583
 Db |||||LeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 719

RESULT 9

ADA14848

ID ADA14848 standard; protein; 798 AA.

XX AC ADA14848;

XX XX

 XX DT 06-NOV-2003 (first entry)
 XX DE Mouse interphotoreceptor matrix component, IPMC, 150 isoform A.

XX

KW mouse; IPMC 150 isoform A; gene therapy;
 KW interphotoreceptor matrix component; IPMC; ocular disorder;
 KW macular degeneration; photoreceptor death; retinal detachment.
 OS Mus sp.
 XX US2002160954-A1.
 XX 31-OCT-2002.
 XX 08-NOV-2001; 2001US-00007270.
 XX 29-OCT-1998; 98US-00183972.
 XX 29-OCT-1999; 99US-00430195.
 XX (IOWA) UNIV IOWA RES FOUND.
 XX PA
 XX Hageman GS, Kuehn MH;
 XX WPI; 2003-238235/23.
 XX N-PSDB; ADA14847.
 XX
 XX New isolated or recombinant interphotoreceptor matrix component
 XX polynucleotide and polypeptide, useful for diagnosing, preventing,
 XX treating or prognosticating ocular disorders, e.g. macular degeneration
 XX or retinal detachment.
 XX
 XX Claim 8; Page 41-43; 76pp; English.
 XX
 XX The invention relates to an isolated or recombinant interphotoreceptor
 XX matrix component (IPMC) polynucleotide. Also disclosed is a vector
 XX comprising a promoter of an interphotoreceptor matrix component (IPMC)
 XX gene operatively linked to the IPMC polynucleotide. The IPMC
 XX polynucleotides, polypeptides and antibodies are useful for diagnosing,
 XX preventing, treating or prognosticating ocular disorders, e.g. macular
 XX degeneration, photoreceptor death or retinal detachment. They are also
 XX useful for identifying a compound capable of modulating IPMC gene
 XX expression in a cell. The present sequence represents the amino acid
 XX sequence of mouse interphotoreceptor matrix component, IPMC, 150 isoform
 XX A.
 XX
 XX SQ Sequence 798 AA;
 Alignment Scores:
 Pred. No.: 1.38e-222 Length: 798
 Score: 2504.50 Matches: 527
 Percent Similarity: 73.13% Conservatives: 80
 Best Local Similarity: 63.49% Mismatches: 180
 Query Match: 42.63% Indels: 43
 DB: 6 Gaps: 10
 US-10-007-270-1 (1-3330) x ADA14848 (1-798)
 QY 131 ATGTATTTGGAACTAGACAGCTATTTTGGTATTTTGGATTTTCCCAAGTTCAAGGA 190
 Db 1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCTCCATTAACATATACCATTTCTGAACTAAGACATAGACATCCOCCA 250
 Db 21 IleLysAspThrSerIleLysIlePheSerSerGluIleLysAsnIleAspLysThrPro 40
 QY 251 AGAATGAAACACTGAAGTACTGAAATAATGACAAATGTCAACTATGAGACGATA 310
 Db 41 ArgIleGluThrIleGluSerThrSerThrValHisLysValSerThrMetLysArgIle 60
 QY 311 TTCCATTGGCAAGCATGCAACAAAGATCGGATTTTCCCAACGGGGGTTAAAGTC 370
 Db 61 PheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro---AlaAlaAsnIle 79
 QY 371 TGTCCAGGAATCCATGAACAGATTTTAGACAGCTTCTCACTTATATAGATTGAGA 430
 Db 80 CysProGlnGluSerLeuArgGlnIleLeuAlaSerLeuGlnIleuTyfTyfArgLeuArg 99

QY 431 GTGTGTGACAGACAGTATGGGAAGCATATCGGATCTTTCTGGATCGCATCCCTGACACA 490
 Db 100 ValCysGlnGluValValTrpGluAlaTrpArgIlePheLeuAspArgIleProAspThr 119
 QY 491 GGGGAATATCAGGACTCGGTGACATCTGCCAGCAGGAGACCTCTGGCTTTTGACATT 550
 Db 120 GluGluTyfGlnAspTrpValSerLeuCysGlnLysGluThrPheCysLeuPheAspIle 139
 QY 551 GGAATAAACTTCAGCAATCCAGGAGACCTGATCTTCCAGCAGAGAAATAAAGAG 610
 Db 140 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnArgIleLysGln 159
 QY 611 AGAAGTTTCCCTGACAGAAAGATGAAATATCTGACAGAGACATTTGGAGAGCTGTGT 670
 Db 160 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 179
 QY 671 GAAACCATTTGTCATTCAACAGCATCTACATTTCAAGACATTTGGGAGCATTTCTAAGA 730
 Db 180 GluAlaProValValPro---Thr----- 186
 QY 731 AACCTTCAGAGAGCAAAATTCAGATGTTGCCAAGCTCTCACTTGGCCCTTTCCCTCTCA 790
 Db 187 -----AspValSerArgMetSerLeuGlyProPheProLeuP 199
 QY 791 CTCCTGATGACACCTCTCAATGAATTCGATAATACACTCAACAGACACCAAGATGC 850
 Db 199 roSerAspAspThrAspLeuLysGluIleLeuSerValThrLeuLysAspIleGlnLysP 219
 QY 851 CTACACAGAAAGAGAAACA-----GAATTCGCTGTGTGGAGAGGC 892
 Db 219 roThrThrGluSerLysThrGluProIleHisValSerGluPheSer-----SerGlu 237
 QY 893 AGAGGTGGAGCTAGCCTCTCTGTGTAACAGAGTTCAGAGGAGAGCTCGCTGACT 952
 Db 237 IuLysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS 257
 QY 953 CCCAGTCCCCATATTACAGAGCTAGCAGAAAGTCCCAACTTCAGATGCAAAAGATAT 1012
 Db 257 erGlySerProLysTrpGlnGluValGlnSerGlnLeuGlnLysIleP 277
 QY 1013 TTAAGAACTTCCAGGATTCAAAATAATCCATGTTGTAGGATTTAGACCAAGAGAAAGA 1072
 Db 277 heLysLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProLysLysGlu 297
 QY 1073 AAGATGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTGCAG 1132
 Db 297 IuAspGlySerSerThrThrGluIleGlnLeuMetAlaIlePheLysArgAspHisAla 317
 QY 1133 AAGCAAAAGCCCTGCAAGTACCTCTCTTTTGTATTCACAAATAATTGAAGTGAGG 1192
 Db 317 IuAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleGluSerGluA 337
 QY 1193 AGCTTCATCATGGAACCATCGAGGAGGACCAAGCAACAGAAATCTATCTCACAGTACAG 1252
 Db 337 rgIleHisGlyValIle---GluAspLysGlnProGluThrTrpLeuThrAlaThrA 356
 QY 1253 ACCTCAAAAGCTGATCAGCAAGACATAGAGGAGAAACAATCTTTGGATGTGGGACAA 1312
 Db 356 spuLysLysLeuIleIleGlnLeuLeuAspGlyAspLysSerLeuValGluGlyLysI 376
 QY 1313 TTCAGTTCACTGATGAATTTGCTGATCAGTCCAGCCCTTTGGTCTCTGACACCAATCAG 1372
 Db 376 LeProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA 394
 QY 1373 AGCTGCCACATCTTTTGTGTTTAAACAGAGATGCTACTTTGAGTCCAGAACTCTCTC 1432
 Db 394 spuLeuProLysProLeuAlaAspValThrLysAspAlaThrLeuSerProGluLeuProp 414
 QY 1433 CTGTTTGAACCCAGCTTGAGACACTGACGAGAGAGACATGGTCTTACCT----- 1482
 Db 414 heValGluProArgGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS 434
 QY 1483 -----GACACTTCTGGTCTCCACCTGCTATGGCCCTCTACTCCTGTCAGAGCTCAC 1537

```

Db 434 erlyAspSerSerProValSerAlaSerIleSerArgSerGluAsnLeuP 454
QY 1538 CTTCTTTATGGCATCAGCATCTTCTCTGACTGATCAAGGCACACAGATACATGG 1597
Db 454 roSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProLeuMet 473
QY 1598 CCACCTGACACAGATAGTAGTACAGGGCTCACATCCACCCAGATGATTCTGCAA 1657
Db 473 hrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTyrSerThrI 493
QY 1658 TCAGCCAACTGGCTCGGAATTTCCATCCACTGCTCATCTTCAGATGACAGCCCATCAA 1717
Db 493 leArgGlnLeuProLeuGluSerSerHisTrpProAlaSerSerSerAspArgGluLeuI 513
QY 1718 GTGCGAGTGGCGAAGATATGTCACACACCTAGATGAATGATCTCTGACACTCTCTG 1777
Db 513 leThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspThrProA 533
QY 1778 CCCCATCTGAGTACACAGCTCAGCAATATGTTCTGTCGCCAGATCAATTTCTGGAGG 1837
Db 533 laeUserGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheLeuGluM 553
QY 1838 ATACCACTCTCTGCTCAGCTTTACAGTATATACCACTAGTCTTATGACCATGCCCCA 1897
Db 553 etThrThrProLeuProThrValArgPheIleThrSerSerGluThrIleAlaThrL 573
QY 1898 AGGCCGAGAGCTGGTAGTGTCTTCAGTCTGCGTGTCTTAACATGGCTTCTTCCAACG 1957
Db 573 ysGlyGlnGluLeuValPhePheSerIleArgValAlaAsnMetProPheSerTyrA 593
QY 1958 ACTGTTTCAACAGAGCTCTCGAGTACCGAGCTCTGAGCACTGAGCAACAATTCACAGCTGC 2017
Db 593 sPLeuPheAsnLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThrAspLeuL 613
QY 2018 TGGTTCCATATCTACGATCCCAATCTTACAGGATTTAAGCACTTGAATTAATTAATTC 2077
Db 613 euValProTyrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluIleLeuSerPheA 633
QY 2078 GAACGGGAGTGTGATTGTGAATAGCAAAATGAAGTTGCTTAAGTCTGCGGTATAACC 2137
Db 633 rGAsnGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProTyrAsnL 653
QY 2138 TCACAGAGGCTGTCACGGGCTCTGAGGATTTGCTCTGTCGAGCCCAACCACTCC 2197
Db 653 euThrGlnAlaValArgGlyValLeuGluAspLeuArgSerThrAlaAlaGlnGlyLeuA 673
QY 2198 ATCTGGAATPAGCAGCTACTCTCTCAACATTGAACCACTGATCAAGCAGATCCCTGCA 2257
Db 673 snLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspProCysL 693
QY 2258 AGTTCCTGGCTGCGCGAATTTGCCATGTGTAAGAACGACGAGTCTGAGGAGCGG 2317
Db 693 ysLeuLeuAspCysGlyLysPheAlaGlnCysValLysAsnGluTyrThrGluGluAlaG 713
QY 2318 AGTGTGCTGCAAAACAGGATATGACAGCCAGGAGCGCTGAGCGCTGGAACACGCGC 2377
Db 713 luCysArgCysArgGlnGlyHisGluSerHisGlyThrLeuAspTyrGlnThrLeuAsnL 733
QY 2378 TCTGTGCTGCTGCAAAAGGAATGCGAGGTCCTCCAGGAAGAGGAGCTCCATGCACT 2437
Db 733 euCysProGly---LysThrCysValAlaGlyArgGluGlnAlaThrProCysArgP 752
QY 2438 TGCCAGATCACTGAAATCAGCATACAAATAGTGTATAAAGTTCCAAATCAAC 2497
Db 752 roProAspHisSerThrAsnGlnAlaGlnGluProGlyValLysLysLeu-----ArgG 770
QY 2498 AAAATAACAGGTAATCAGTAAAGAAATTTCTGAATTAATCTGACCTGAGATATCAAGAT 2557
Db 770 lnGlnAsnLysValValLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluP 790
QY 2558 TTAACCATCAAGATTGGAAGAAAT 2583

```

Db 790 heGluAspGlnAspTrpGluGlyAsn 798

RESULT 10

ABR42345

XX ABR42345 standard; protein; 798 AA.

XX ABR42345;

XX 11-AUG-2003 (first entry)

XX Mouse interphotoreceptor matrix IPM 150, isoform A.

XX Mouse; interphotoreceptor matrix; IPM 150; IPMC; receptor; ophthalmological; gene therapy.

XX Mus sp.

| Key | Location/Qualifiers |
|--------|-----------------------------|
| Domain | 71..80 |
| Domain | /note= "N-terminal domain" |
| Domain | 95..125 |
| Domain | /note= "conserved domain 1" |
| Domain | 405..577 |
| Domain | /note= "mucin-like domain" |
| Domain | 576..657 |
| Domain | /note= "conserved domain 2" |
| Domain | 697..740 |
| Domain | /note= "EGF-like domain" |

XX WO2003039346-A2.

XX 15-MAY-2003.

XX 08-NOV-2002; 2002WO-US036090.

XX 08-NOV-2001; 2001US-00077270.

XX (IOWA) UNIV IOWA RES FOUND.

XX Hageman GS, Kuehn MH;

XX WPI; 2003-441440/41.

XX N-PSDB; ACC57950.

XX New interphotoreceptor matrix proteins and polynucleotides, useful for treating or preventing photoreceptor death or retinal detachment, or for treating ocular disorders.

XX Claim 6; Page 82; 105pp; English.

The present sequence is the protein sequence of isoform A of novel mouse interphotoreceptor matrix 150 (IPM 150), a member of the newly identified interphotoreceptor matrix component (IPMC) gene family. IPMC gene family members have been identified in humans, monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2 Subfamilies are designated IPM 150 (or IPM1) and IPM 200 (or IPM2). The invention provides IPM 150 and IPM 200 polynucleotides and polypeptides, antibodies that specifically bind the polynucleotides, and vectors comprising the polynucleotides. A claimed method of treating or preventing photoreceptor death or retinal detachment involves administering an IPMC polynucleotide, polypeptide or antibody. Also claimed is a method for identifying a compound capable of modulating IPMC gene expression

SQ Sequence 798 AA;

Alignment Scores:

| Pred. No.: | 1.38e-222 | Length: | 798 |
|------------------------|-----------|---------------|-----|
| Score: | 2504.50 | Matches: | 527 |
| Percent Similarity: | 73.13% | Conservative: | 80 |
| Best Local Similarity: | 63.49% | Mismatches: | 180 |
| Query Match: | 42.63% | Indels: | 43 |
| DB: | 7 | Gaps: | 10 |

| | | | |
|----|------|--|------|
| QY | 1193 | AAGTCTATCATGAACCATGGAGAGCAGCAGCAACAGAAATCTATCTCTCAGCTACG | 1255 |
| Db | 337 | rgflleHisHisGlyValIle:::---GluAspLysGlnProGluThrTyrLeuThrAlaThrA | 356 |
| QY | 1253 | ACCTCAAAAGGCTGATCAGCAAAAGCATAGAGGAAGAACAACTCTTTGGATGTGGGACAA | 1312 |
| Db | 356 | spLeuLysLysLeuIleIleGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI | 376 |
| QY | 1313 | TTCAAGTTCTACTGATGAATTTCTGGATGATCTGCAGCGCTTTTGGTCTGTGACACCAATCAG | 1372 |
| Db | 376 | leProPheGlyAspGluValThrGlyThrLeu:::---PheArgProValThrGluProA | 394 |
| QY | 1373 | AGCTGGCCACATCTTTTCTGTTTATAACAGAGGATGCTCTTTAGTCCAGACTTCCTC | 1432 |
| Db | 394 | spLeuProLysProLeuAlaAspValIthrGluAspAlaThrLeuSerProGluLeuProP | 414 |
| QY | 1433 | CTGTTGAACCCGAGCTTGAGACAGTGGAGCGGACAGCATGGTCTACCT:::----- | 1482 |
| Db | 414 | heValGluProArgGlnGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS | 434 |
| QY | 1483 | -----CACACTTCTTGGTCTCCACCTGCTATAGGCTCTTACCTCCTGTCAGAGCTCCAC | 1537 |
| Db | 434 | erLysAspSerSerTrpSerProProValSerAlaSerIleSerArgSerGluAsnLeuP | 454 |
| QY | 1538 | CTTTCTTTATGGCATCAAGCATCTCTCTCTCTGACTGATCAAGCCACACAGATACAATGG | 1597 |
| Db | 454 | roSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProLeuMetI | 473 |
| QY | 1598 | CCACTGACCAAGACATCTAGTACAGGGCTCACCTATAGGCTCTTACCTCCTGTCAGAGCTCCAC | 1657 |
| Db | 473 | hrThrGlyProThrAlaLeuIleProLysProThrLeuLeuProThrIleAspLysSerThrI | 493 |
| QY | 1658 | TCAGCCAACTGGCTCTGGGAATTTCCACATCCACCTCATCTTTCAGATGACGACCGCATCAA | 1717 |
| Db | 493 | leArgGlnLeuProLeuGluSerSerHisTrpProIlaSerSerSerAspArgGluLeuI | 513 |
| QY | 1718 | GTGCAAGTGGCGAAGATATGGTCACACACCTTAGATGAATGGATCTCTGTGACACTCCCTG | 1777 |
| Db | 513 | leThrSerSerHisAspThrIleAsgAspLeuAspGlyMetAspValSerAspThrProA | 533 |
| QY | 1778 | CCCATCTGAGTACACAGACTCAGCAATATGTTCTGTCCAGATCATCTTCTGGAGG | 1837 |
| Db | 533 | laLeuSerGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheLeuGluM | 553 |
| QY | 1838 | ATACCACTCTGTCTCAGCTTTACAGTATATACCACTAGTTCTTATGACCATGGCCCCA | 1897 |
| Db | 553 | euThrThrProIleProThrValArgPheIleThrThrSerSerGluThrIleAlaThrL | 573 |
| QY | 1898 | AGGGCCGAGAGCTGTTAGTGTCTTCTAGTCTGGGTGTGCTTAAACATGGCCCTTCTCAAAG | 1957 |
| Db | 573 | YsGlyGlnGluLeuValPhePheSerLeuArgValAlaAsnMetProPheSerTyrA | 593 |
| QY | 1958 | ACCTGTTCAACAAGAGCTCTCTGGAGTACCGAGCTCTGAGCAACAATTCACACAGCTGC | 2017 |
| Db | 593 | spLeuPheAsnLysSerSerLeuGluTyrGlnAlaLeuGluArgPheThrAspLeuI | 613 |
| QY | 2018 | TGTTTCCTATCTACGATCCCAATCTTACAGATTTAAGCAACTTGAATACTTAATCTCA | 2077 |
| Db | 613 | euValProTyrLeuAArgSerAsnLeuThrGlyPheLysGlnLeuGluLeuSerPheA | 633 |
| QY | 2078 | GAAACGGGAGTGTGTTGTAATAGCAAAATGAGCTTTGCTTAGTCTGTGGCGGTATAACC | 2137 |
| Db | 633 | rgAsnGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProTyrAsnL | 653 |
| QY | 2138 | TCACCAAGGCTGTGCACGGGCTCTTGGAGGATTTTCGTTCTGTCTGCAGCCCAACATCC | 2197 |
| Db | 653 | euThrGlnAlaValAArgGlyValLeuGluAspLeuArgSerThrAlaAlaGlnGlyLeuA | 673 |
| QY | 2198 | ATCTGGAAATAGACAGCTACTCTCTCAACATTTGAACACAGCTGTATCAAGCAGATCCCTGCA | 2257 |
| Db | 673 | snLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspProCysL | 693 |

Db 146 erGlulysValGluPheSerIleSerLeuProIleHisArgPheIleAlaGluLeuT 166
 QY 947 CTGACTCCCACTCCCATATTACAGGAGCTAGCAGAAAGTCCCACTTCAGATGCAAA 1006
 Db 166 hrAsnSerGlySerProTyrTyrGlnGluLeuValGlyGlnSerGlnGluGln 186
 QY 1007 AGATATTTAAGAACTCCAGGATTCAAAATAATCATGTGTAGGATTTAGACCAAGA 1066
 Db 186 yIlePheIleGlyLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProIle 206
 QY 1067 AAGAAAGATGGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGAGACACA 1126
 Db 206 yslGluAspGlySerSerSerThrGlnIleGlnLeuMetAlaIlePheIleYsArgSPH 226
 QY 1127 GTGCGAAGCAAAAGCCCTCGAAGTCTCTCTGTTTGTATTCACAAATAATGAAA 1186
 Db 226 iAlaGluAlaIleYsSerProAspSerHisLeuLeuSerLeuAspSerAsnIleGluS 246
 QY 1187 GTGAGGAAGTCTATCATGCAACCATGAGGAGGACACCAACCAACCAATCTATCTCAAG 1246
 Db 246 erGluArgIleHisIleGlyValIle---GluAspIleGlnProGluThrTyrLeuThra 265
 QY 1247 CTACAGACTCAAAAGGCTGATCGAAGCAACTAGAGGAAGAAACAATCTTTGGATGG 1306
 Db 265 laThrAspLeuIleYsLeuIleIleGlnLeuLeuAspIleYsPleuSerLeuValGluG 285
 QY 1307 GGACAATTCAGTCTCATGATGAATTCGTATCACTCCAGCCTTTGGTCTCGACACC 1366
 Db 285 lYleIleProPheGlyAspGluValThrGlyThrIleu-----PheArgProValThrG 303
 QY 1367 AATCAGAGCTGCCACATCTTTTGTCTGTATACAGAGATGCTACTTTGAGTCCAGAAC 1426
 Db 303 luProAspLeuProIysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluL 323
 QY 1427 TTCCTCTGTTGAACCCAGCTTGAGACAGTGGAGCGGAGCAGACATGCTCTACT--- 1482
 Db 323 euProPheValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuProGlyM 343
 QY 1483 -----GACACTTCTGCTCCAGCTGCTATGGCTCTTACTCTCTCCCTGTCAGAG 1531
 Db 343 etSerSerIysAspSerTrpSerProValSerAlaSerIleSerArgSerGluA 363
 QY 1532 CTCACACTTCTTTATGGCATCAGCATCTTCTCTCTACTGATCAGCAGCCACACAGATA 1591
 Db 363 snLeuProSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProL 382
 QY 1592 CAATGGCCACTGACAGCAATGCTAGTACAGGCTCACCATCCCAACACAGTATTATT 1651
 Db 382 euMetThrThrGlyProThrAlaLeuIleProIysProThrLeuProThrIleAspTyrS 402
 QY 1652 CTGCAATCAGCCAACTGGCTCTGGAAATTCATCCACCTGATCTTCAGATCAGACAGCC 1711
 Db 402 erThrIleArgGlnLeuProLeuGluSerSerHisTrpProAlaSerSerArgG 422
 QY 1712 GATCAAGTGCAGTGGCAAGATATGTCAGACACTAGATGAATGATCTGCTCACA 1771
 Db 422 luLeuIleThrSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspT 442
 QY 1772 CTCCTGCCCCATCTGAGTACAGAGCTCAGCGAATATGTTTCTGCCCCAGATCATTTCT 1831
 Db 442 hrProAlaLeuSerGluIleSerGluLeuSerGlyTyrAspSerAlaSerGlyGlnPheL 462
 QY 1832 TGGAGGATACCACTCTCTCAGCTTTTACAGTATACACACTAGTTCTATGACCATG 1891
 Db 462 euGluMetThrThrProIleProThrValArgPheIleThrThrSerSerGluThrIleA 482
 QY 1892 CCCCAGGCGCGAGAGCTGTGTGTCTTCAGTCTGCTGTTGTGTAACATGCGCTTCT 1951
 Db 482 laThrIleYsGlyGlnGluLeuValIlePhePheSerLeuArgValAlaAsnMetProPheS 502
 QY 1952 CCAACGACTGTTTCAACAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACAC 2011
 Db 502 erTyrAspLeuPheAsnIysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThra 522

QY 2012 AGCTGTGTTTCCATATCTAGCATCCATCTTACAGATTTAAGCAACTTCAAACTACTTA 2071
 Db 522 splLeuLeuValProTyrIleuArgSerAsnLeuThrGlyPheIysGlnLeuGluIleLeuS 542
 QY 2072 ACTTCAGAAACGGGAGTGTGATTGTGAATAGCAAAATGAAGTTTCTAAGTCTGTGCGCT 2131
 Db 542 erPheArgAsnGlySerValIleValenSerIysValargPheAlaIysAlaValProt 562
 QY 2132 ATAACCTCACCAAGCGTCTCCACGGGTCTTGGAGGATTTTGTCTCTCTGCGAGCCCAAC 2191
 Db 562 yzAsnLeuThrGlnAlaValargGlyValLeuGluAspLeuArgSerThrAlaAlaGlnG 582
 QY 2192 AACTCCATCTCGAAATAGACAGCTACTCTCTCAACATTGAACACCTGATCAAGCAGATC 2251
 Db 582 lYleAsnLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspP 602
 QY 2252 CTGCGAAGTCTCTGCGCTCGCGGAAATTGCCAATGTGTAAAGAACCAACGAGCTGAGG 2311
 Db 602 roCylsIysLeuLeuAspCysGlyLysPheAlaGlnCysValIysAsnGluTrpThrGluG 622
 QY 2312 AAGCGAGCTCTGCTGCAAAACAGGATATGACACCCAGGGAGCTCGACGCTTGGAAAC 2371
 Db 622 luAlaGluCysArgCysArgGlnGlyHisGluSerHisGlyThrLeuAspTyrGlnThrL 642
 QY 2372 CAGGCTCTGTGCGCTCGCACAAAGGAATGCGAGGTCTCCACGGGAAGGGAGCTCCAT 2431
 Db 642 euAsnLeuCysProProGly---LysThrCysValIleAlaGlyArgGluGlnAlaThrProC 661
 QY 2432 GCAGGTGCGAGTACTCTGAAATCAAGCATACAAAATAGTGTTTAAAGAGTTCCTCAA 2491
 Db 661 ysArgProAspHisSerThrAsnGlnAlaGlnGluProGlyValIysIysLeu---- 679
 QY 2492 ATCAACAAATAACAGTAATCAGTAAAGAAATTCAGTAAAGAAATTCAGTAAAGATG 2551
 Db 680 --ArgGlnGlnAsnLysValIysLysArgAsnSerIysLeuSerAlaIleGlyPheG 699
 QY 2552 AAGAAATTTAACCATCAAGATTGGGAAGCAAT 2583
 Db 699 luGluPheGlu**GlnAspTrpGluGlyAsn 709
 RESULT 12
 ID ADA14850 standard; protein; 466 aa.
 AC ADA14850;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Mouse interphotoreceptor matrix component, IPMC, 150 isoform D.
 XX mouse; IPMC 150 isoform D; gene therapy;
 KW interphotoreceptor matrix component; IPMC; ocular disorder;
 KW macular degeneration; photoreceptor death; retinal detachment.
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT Protein /note= "signal sequence"
 FT 21..466
 FT /note= "Mature IPMC 150 isoform D"
 XX US2002160954-A1.
 XX 31-OCT-2002.
 XX 08-NOV-2001; 2001US-00007270.
 XX 29-OCT-1998; 98US-00183972.
 XX 29-OCT-1999; 99US-00430195.
 XX (IOWA) UNIV IOWA RES FOUND.

XX Hageman GS, Kuehn MH;
 XX WP; 2003-238235/23.
 DR N-PSDB; ADAL4849.
 XX New isolated or recombinant interphotoreceptor matrix component
 PT polynucleotide and polypeptide, useful for diagnosing, preventing,
 PT treating or prognosticating ocular disorders, e.g. macular degeneration
 PT or retinal detachment.
 XX Claim 8; Page 44-45; 76pp; English.
 XX The invention relates to an isolated or recombinant interphotoreceptor
 CC matrix component (IPMC) polynucleotide. Also disclosed is a vector
 CC comprising a promoter of an interphotoreceptor matrix component (IPMC)
 CC gene operatively linked to the IPMC polynucleotide. The IPMC
 CC polynucleotides, polypeptides and antibodies are useful for diagnosing,
 CC preventing, treating or prognosticating ocular disorders, e.g. macular
 CC degeneration, photoreceptor death or retinal detachment. They are also
 CC useful for identifying a compound capable of modulating IPMC gene
 CC expression in a cell. The present sequence represents the amino acid
 CC sequence of mouse interphotoreceptor matrix component, IPMC, 150 isoform
 CC D.
 XX Sequence 466 AA;
 SQ

Alignment Scores:
 Pred. No.: 7,73e-111 Length: 466
 Score: 1298.50 Matches: 309
 Percent Similarity: 43.76% Conservative: 52
 Best Local Similarity: 37.45% Mismatches: 99
 Query Match: 22.10% Indels: 365
 DB: 6 Gaps: 10

US-10-007-270-1 (1-3330) x ADAL4850 (1-466)

QY 131 ATGTATTTGGAACTAGAGAGCTATTTTGGATTTTCCAGTTCAGGA 190
 DB 1 MetAenPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAAGATATCTCCATTAACATATACCTCTGAAACTTAAAGACATAGACATCCCCA 250
 DB 20 ----- 20
 QY 251 AGAATGAAACAACCTGAAAGTACTGAAAAAATGTCAMAAATGTCATATGAGACGAATA 310
 DB 20 ----- 20
 QY 311 TTCGATTGGCAAGCATCGAACAAAAGATCCGCATTTTCCCAACGGGGTTAAAGTC 370
 DB 20 ----- 20
 QY 371 TGTCCACAGGAATCCATGAAACAGATTTTACAGCTTTCAGAGCTTATATAGATTGAGA 430
 DB 21 -----11leys 22
 QY 431 GTGTCTCAGGAGCAGTATGGAAGCATATCGCATTTCTGGATCCCATCCCTGACACA 490
 DB 23 ValCysGlnGluValTrpGluAlaTrpArgIlePheLeuAspArgIleProAspThr 42
 QY 491 GGGGAATATCAGGACTGGGTGAGCATCTGCAGCAGGAGACTTCTGCTCTTTGACATT 550
 DB 43 GluGluTyGlnAspTrpValSerLeuCysGlnLysGluThrPheCysLeuPheAspIle 62
 QY 551 GGAATAATCTCAGCAATTCCTCAGGACCTGATCTTCTCCAGCAGAGAAATAAAGACAG 610
 DB 63 GlyLysAsnPheSerAenSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 82
 QY 611 AGAGTTTCCCTGACAGAAAAGATGAATATCTGCAGAGAGACATTTGGGAGGCTGGT 670
 DB 83 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 102

QY 671 GAAACCAATGTCTATTTCAACAGCAATCTACATTTTCAAGAGCTTGGGCGAGTATTTCAAGAA 730
 DB 103 GluAlaProValPro--Thr----- 109
 QY 731 AACCTCAGAGAGCAAAATTCAGATGTTGCCAAGCTCTCACTTGGGCTTCCCTCTCA 790
 DB 110 -----AspValSerArgMetSerLeuGlyProPheLeuP 122
 QY 791 CTCTGTATGACACCTCTCAATGAAATTCGATATATACACTCAACACCAACCAAGATCC 850
 DB 122 roSerAspAspThrAspLeuLysGluLeuSerValThrLeuLysAspIleGlnLysP 142
 QY 851 CTACACAGAAAGAAACA-----GAATTCCTGTGTGGAGAGC 892
 DB 142 roThrThrGluSerIleThrGluProIleHisValSerGluPheSer-----SerGluG 160
 QY 893 AGAGGTGTCAGCTCAGGTCTCTCTGTGTAACACAGAGTTCAAGCAGCAGCTCGCTGACT 952
 DB 160 luySValGluPheSerIleSerLeuProAenHisArgPheLysAlaGluLeuThrAsnS 180
 QY 953 CCAGTCCCATATTTACCAGGAGCTAGCAGGAAAGTCCCAACTTCAGATGCAGAAATAT 1012
 DB 180 ergLysSerProTy-TyrGlnGluLeuValGlyGlnSerGlnLeuGlnLeuGlnLysP 200
 QY 1013 TTAAGAACTTCCAGGATTCAAAAAATCCATGTGTAGATTAGACCAAAAGAAAGAA 1072
 DB 200 helYsLeuProGlyPheGlyIleArgValLeuGlyPheArgProLysLysGluG 220
 QY 1073 AAGATGGCTCAAGCTCCACAGAGATGCAACTTACGGCCATCTTTAAGACACACAGTGCAG 1132
 DB 220 luyAspGlySerSerThrGluLeuGlnLeuMetAlaIlePheLysArgAspHisAlaG 240
 QY 1133 AGCAAAAAGCCCTGCAAGTGACCTCTGTCTTTGATTCCCAACAAATTTGAAGTGAGG 1192
 DB 240 luAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleLysSerGluA 260
 QY 1193 AGCTCTATCATGGAACCATGAGGAGGACAGCAACATCTTTGGATGTGGGACAA 1312
 DB 260 rgIleHisGlyValIle---GluAspLysGlnProGluThrTyrLeuThrAlaThrA 279
 QY 1253 ACCTCAAAAGGCTGATCAAGAACCATAGAGGAAGAACAACTTTTGGATGTGGGACAA 1312
 DB 279 spleuLysLysLeuIleLeuGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI 299
 QY 1313 TCCAGTTCATGATGAAATTCGTGATCATCTGCCAGCCTTTGGTCTGACACCAATCAG 1372
 DB 299 leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA 317
 QY 1373 AGCTGCCACATCTTTTGTGTTTATAACAGAGGATGCTACTTTGAGTCCAGAACTTCTTC 1432
 DB 317 spleuProLysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluLeuProP 337
 QY 1433 CTGTTGAACCCAGCTTTGAGACAGTGGACGAGGAGCAGAGCATGGTCTACCTGACACTTCTT 1492
 DB 337 heValGluProArgLeuGluAlaValAsp----- 346
 QY 1493 GGTCTCCACCTGTATGGGCTCTACCTCCCTGTGACAGCTCCACCTTTCTTTATGGCAT 1552
 DB 346 ----- 346
 QY 1553 CAAGCATCTTCTCTGACTGATCAAGGCACACACAGATACATACATGCCCACCTGACACACA 1612
 DB 346 ----- 346
 QY 1613 TGCTAGTACAGGCTCACCATCCCCACAGTGTATTTCTGCAATCAGCCAACCTGGCTC 1672
 DB 346 ----- 346
 QY 1673 TGGGAATTTCACTCCACTGCACTTCAGATGACAGCGCATCAAGTGCAGGTGGCGAAG 1732
 DB 346 ----- 346
 QY 1733 ATATGTCAGACACCTAGATGAATGGATCTGTCTGACACTCTCTGCCCCATCTGAGGTAC 1792

Db 346 ----- 346
QY 1793 CAGAGCTCAGGAATATGTTCTGTCCAGATCAATTTCTTGAGGATACCACTCTCTGCT 1852
Db 346 ----- 346
QY 1853 CAGCTTTACAGATATATCACCAGTATGTTCTATGACCATTTGCCCCCAAGGCGGAGAGCTGG 1912
Db 346 ----- 346
QY 1913 TAGTGTCTTTCAGTCTGCGTGTCTAAGTCTGCTTAACATGCGCTTCTCCACGACCTGTTCAACAAGA 1972
Db 346 ----- 346
QY 1973 GCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACAGCTGCTGTTCCATATCTAC 2032
Db 346 ----- 346
QY 2033 GATCCAACTTTACAGGATTAAGCAACTTGAATACTTAATTCAGAAACGGAGTGTGA 2092
Db 347 ----- 352
QY 2093 TTGTGAATAGCAAAATGAAGTTTCTAAGTCTGTGCGGTATAACCTCACCAGGCTGTGC 2152
Db 352 eu----- 352
QY 2153 ACGGGTCTTGGAGGATTTTGTCTGTGCGAGCCCAACAACTCCATCTCGAAATAGACA 2212
Db 352 ----- 352
QY 2213 GCTACTCTCTCAACATTGAACAGCTGATCAAGCAGATCCCTGCAAGTCTCTGGCTGG 2272
Db 353 ----- 366
QY 2273 GCGAATTTGCCATGCTGTAAGAACGACGAGTCTGAGAGCGGAGTCTGCTGCAAC 2332
Db 366 llyyepheAlaGlnCysVallysaenGlnTrpThrGluGluAlaGlnCysArgCysArgG 386
QY 2333 CAGGATATGACAGCGGAGCGCTGAGCGCTTGGACCGCTCTGTGGCCCTGGCA 2392
Db 386 lnglyHisGluSerHisGlyThrLeuAspTyGlnThrLeuAsnLeuCysProProGly- 405
QY 2393 CAAGGAATGCGAGGTCTCCAGGAAAGGAGTCTCCATGAGTTCAGATTCATCTCTG 2452
Db 406 --LysThrCysValAlaGlyArgGluGlnAlaThrProCysArgProThrAspHisSerT 425
QY 2453 AAAATCAAGCATACAACTAGTGTGTAAGTTCCAAATCAACAAATACAGGTAA 2512
Db 425 hrAsnGlnAlaGlnGluProGlyVallyLysLeu-----ArgGlnGlnAsnLysValV 443
QY 2513 TCAGTAAAGAAATTTCTGAATTAATCTGACCGTGAATATGAAAGATTTAAACCATCAAGATT 2572
Db 443 alLysLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluPheGluAspGlnAspT 463
QY 2573 GCGAGGAAT 2583
Db 463 rpGluGlyAsn 466
RESULT 13
ABR42346
XX ABR42346 standard; protein; 466 AA.
XX AC ABR42346;
XX 11-AUG-2003 (first entry)
XX DE Mouse interphotoreceptor matrix IPM 150, isoform D.
XX KW Mouse; interphotoreceptor matrix; IPM 150; IPMC; receptor;
XX KW ophthalmological; gene therapy.
XX OS Mus sp.

XX Key Location/Qualifiers
FH Peptide 1..20
FT /label= Signal_peptide
FT Protein 21..466
FT /label= IPM_150
FT Domain 21..49
FT /note= "conserved domain"
FT Domain 360..408
FT /note= "EGF-like domain"
XX WO2003039346-A2.
PN 15-MAY-2003.
PD
XX 08-NOV-2002; 2002WO-US036090.
PF
XX 08-NOV-2001; 2001US-00077270.
PR
XX (IOWA) UNIV IOWA RES FOUND.
PA Hageman GS, Kuehn MH;
XX
PI WPI; 2003-441440/41.
XX N-PSDB; ACG57951.
DR
XX New interphotoreceptor matrix proteins and polynucleotides, useful for
PT treating or preventing photoreceptor death or retinal detachment, or for
PT treating ocular disorders.
XX
PS Claim 6; Page 83; 105pp; English.
XX
CC The present sequence is the protein sequence of isoform D of novel mouse
CC interphotoreceptor matrix 150 (IPM 150), a member of the newly identified
CC interphotoreceptor matrix component (IPMC) gene family. IPMC gene family
CC members have been identified in humans, monkey, cow, goat, rabbit, dog,
CC cat, pig, mouse and rat IPM. 2 Subfamilies are designated IPM 150 (or
CC IPM61) and IPM 200 (or IPM62). The invention provides IPM 150 and IPM 200
CC polynucleotides and polypeptides, antibodies that specifically bind the
CC polypeptides, and vectors comprising the polynucleotides. A claimed
CC method of treating or preventing photoreceptor death or retinal
CC detachment involves administering an IPMC polynucleotide, polypeptide or
CC antibody. Also claimed is a method for identifying a compound capable of
CC modulating IPMC gene expression
XX
SQ Sequence 466 AA;
Alignment Scores:
Pred. No.: 7,73e-111 Length: 466
Score: 1298.50 Matches: 309
Percent Similarity: 43.76% Conservative: 52
Best Local Similarity: 37.45% Mismatches: 99
Query Match: 22.10% Indels: 365
DB: 7 Gaps: 10
US-10-007-270-1 (1-3330) x ABR42346 (1-466)
QY 131 ATGTATTTGGAAACTAGAGAGCTATTTTGTGTTTGGATTTTCTCCAGTTCAGGA 190
Db 1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20
QY 191 ACCAAAGATATCTCCATTAACATATACCATCTCTGAAACTAAAGACATAGACAAATCCCA 250
Db 20 ----- 20
QY 251 AGAAATGAACAACATGAAAGTACTGAAAAATCTACAAAATGTCAACTATGAGACGAATA 310
Db 20 ----- 20
QY 311 TTCGATTTGGCAAGCATCGAACAAAAAGATCGCATTTTCCCAACGGGGTTAAAGTC 370
Db 20 ----- 20

| | | | | | | | | | | | |
|----|------|-------|------------|----------|---------|----------|---------|---------|--------|--------|------|
| Qy | 371 | TGTC | CCAGGAATCC | AGAAACAG | ATTTTAG | CAGCTTC | CAAGCTT | ATTATAG | ATTAG | ATTCAG | 430 |
| Db | 21 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 431 |
| Qy | 431 | GTG | GTCCAGG | AGCAGT | ATGGG | ACATATCC | GAATCTT | CTTGAT | CGATCC | CTCGAC | 490 |
| Db | 23 | Val | Cy | eGln | GuVal | Val | Trp | Gu | Ala | Trp | 42 |
| Qy | 491 | GGGG | AATATC | AGGATCG | GGT | CAGCAT | CTGC | CACG | AGAC | CTTCTG | 550 |
| Db | 43 | Glu | Gu | Tyr | Gln | Asp | Trp | Val | Ser | Leu | 52 |
| Qy | 551 | GG | AAAACT | TCCAG | ATCC | CAG | AGC | ACTTGA | ATCTT | CCAG | 610 |
| Db | 63 | Gly | Leu | Asn | Phe | Ser | Asn | Ser | Gln | His | 82 |
| Qy | 611 | AGA | AGTTTCC | TGAC | AGAAAG | ATGA | ATATCT | GCAG | AGAG | AGAC | 670 |
| Db | 83 | Arg | Ser | Phe | Pro | Gly | Arg | Val | Asp | Leu | 102 |
| Qy | 671 | GA | ACCAT | TGTC | ATTC | CA | AC | CAATCT | CA | ATTTCA | 730 |
| Db | 103 | Glu | Ala | Pro | Val | Pro | Thr | ----- | ----- | ----- | 109 |
| Qy | 731 | A | ACCTC | AGAG | AGAA | ATTC | AG | ATTTG | CC | AC | 790 |
| Db | 110 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 122 |
| Qy | 791 | CTC | CTGAT | GAC | ACCT | CTCA | TGA | ATCT | CGA | TAATCA | 850 |
| Db | 122 | ro | Ser | Asp | Asp | Thr | Asp | Leu | Val | Thr | 142 |
| Qy | 851 | CTA | CA | AGAA | AGAA | CA | CA | ----- | ----- | ----- | 892 |
| Db | 142 | ro | Thr | Thr | Thr | Thr | Thr | Pro | Leu | His | 160 |
| Qy | 893 | AG | AGGT | CG | AGCT | CG | CTCT | CGT | TA | AC | 952 |
| Db | 160 | lu | Val | Gu | Leu | Phe | Ser | Leu | Pro | Asn | 180 |
| Qy | 953 | CC | CAGT | CCCC | ATAT | TAC | CAG | AGCT | AG | CA | 1012 |
| Db | 180 | er | Gly | Ser | Pro | Tyr | Tyr | Gln | Leu | Val | 200 |
| Qy | 1013 | TTA | AGAA | ACTT | CC | CAG | ATTC | AAAA | ATCC | ATGT | 1072 |
| Db | 200 | he | Leu | Val | Ser | Pro | Gly | Phe | Gly | Val | 220 |
| Qy | 1073 | AG | ATGG | CT | CA | AGT | TC | CA | AGAT | TC | 1132 |
| Db | 220 | lu | Asp | Gly | Ser | Ser | Thr | Gu | Leu | Met | 240 |
| Qy | 1133 | AG | CA | AAAA | AG | CCCT | GC | AGT | CA | CTCT | 1192 |
| Db | 240 | lu | Ala | Val | Ser | Pro | Asp | Ser | His | Leu | 260 |
| Qy | 1193 | AG | CT | AT | CA | TGA | CCAT | CG | AGG | AGG | 1252 |
| Db | 260 | rg | Leu | His | eGly | Val | Leu | ----- | ----- | ----- | 279 |
| Qy | 1253 | AC | CT | AAA | AGG | CTG | AT | CA | AG | AG | 1312 |
| Db | 279 | sp | Leu | Val | Leu | Leu | Leu | Asp | Gly | Asp | 299 |
| Qy | 1313 | TT | CAGT | TC | CA | TG | ATGA | ATCT | GGT | CT | 1372 |
| Db | 299 | le | Pro | Phe | Gly | Asp | Gu | Val | Thr | Gly | 317 |
| Qy | 1373 | AG | CT | CCCC | CA | CAT | CTTT | TG | CT | TA | 1432 |
| Db | 317 | sp | Leu | Pro | Val | Pro | Leu | Ala | Asp | Val | 337 |
| Qy | 1433 | CT | GT | TTGA | AC | CC | AG | CTT | G | AG | 1492 |

[illegible]

/note= "conserved domain"

FT XX WO2003039346-A2.
 XX PN 15-MAY-2003.
 XX PD 08-NOV-2002; 2002WO-US036090.
 XX PF 08-NOV-2001; 2001US-00077270.
 XX PR (IOWA) UNIV IOWA RES FOUND.
 XX PA Hageman GS, Kuehn MH;
 XX PI WPI; 2003-441440/41.
 XX DR N-PSDB; ACC57948.
 XX DR
 XX PT New interphotoreceptor matrix proteins and polynucleotides, useful for
 PT treating or preventing photoreceptor death or retinal detachment, or for
 PT treating ocular disorders.
 XX PS
 XX PS Claim 6; Page 80; 105pp; English.
 XX CC The present sequence is the protein sequence of isoform C of novel human
 CC interphotoreceptor matrix 150 (IPM 150), a member of the newly identified
 CC interphotoreceptor matrix component (IPMC) gene family. The IPM 150 gene
 CC is located on chromosome 6q13-q15, a region that also contains loci for
 CC progressive bifocal choriorretinal atrophy, autosomal dominant Stargard's
 CC -like macular dystrophy, North Carolina macular dystrophy and Salla
 CC disease. Members of the IPMC gene family have been identified in humans,
 CC monkey, cow, goat, rabbit, dog, cat, pig, mouse and rat IPM. 2
 CC Subfamilies are designated IPM 150 (or IPM1) and IPM 200 (or IPM2). The
 CC invention provides IPM 150 and IPM 200 polynucleotides and polypeptides,
 CC antibodies that specifically bind the polypeptides, and vectors
 CC comprising the polynucleotides. A claimed method of treating or
 CC preventing photoreceptor death or retinal detachment involves
 CC administering an IPMC polynucleotide, polypeptide or antibody. Also
 CC claimed is a method for identifying a compound capable of modulating IPMC
 CC gene expression
 XX SQ Sequence 198 AA;

Alignment Scores:
 Pred. NO.: 6.82e-87 Length: 198
 Score: 1038.00 Matches: 197
 Percent Similarity: 99.49% Conservative: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 17.67% Indels: 0
 DB: 7 Gaps: 0

US-10-007-270-1 (1-3330) x ABR42344 (1-198)

QY 131 ATGATTTGGAACTAGAGAGCTATTTTGTGTTTGGATTTTCTCAAGTTCAAGGA 190
 Db 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCTCCATTACATATACATCTGAAATCTAAAGACATAGACAATCCCCCA 250
 Db 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAspAsnProPro 40
 QY 251 AGAAATGAACAACACTGAAAGTACTGAAATCTCAAAATCTCAACTATGAGACGAATA 310
 Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgIle 60
 QY 311 TTGATTTGGCAAGCATCGAACAAGATCGGCATTTTCCCAACGGGGTTAAAGTC 370
 Db 61 PheAspLeuAlaLysHisArgThrIysArgSerAlaPhePheProThrGlyValVal 80
 QY 371 TGTCACAGGAATCCATGAAACAGATTTTAGACAGTCTTCAAGCTTATTATATAGATTGAGA 430
 Db 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100
 QY 431 GTGTGTGAGGAAGCAGATGATGGAGCATATCCGATCTTCTGGATCGATCCCTCGACACA 490

Db 101 ValCysGlnGluAlaAlaTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120
 QY 491 GGGGATATCAGGACTGGGTCTGAGCATCTGCGAGAGAGACCTTCTGCTCTTTGACATT 550
 Db 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140
 QY 551 GGAATAAACTTCAGCAATTCAGGAGGACCTGCTCTCCAGCAGAGAAATAAACACAG 610
 Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnIleArgIleLysGln 160
 QY 611 AGAAGTTTCCCTGACAGAAAAAGATGAAATATCTGACAGAGACATTTGGGAGAGCCTGGT 670
 Db 161 ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGlyGluProGly 180
 QY 671 GAAACCATTTGTCATTTCACAGCAATCTACATTTCAAGACTTGGGAGATTTTC 724
 Db 181 GluThrIleValIleSerThrAlaIleTyrIleSerLysThrTrpAlaValPhe 198

Search completed: March 4, 2004, 18:45:30
 Job time : 198.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 4, 2004, 18:39:36 ; Search time 34 Seconds
(without alignments)
10112.617 Million cell updates/sec

Title: US-10-007-270-1
Perfect score: 5875
Sequence: 1 taaacacgaagggtatcct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DBP=xlp
-Q=/cgn2_1/USPTO.spool_p/US10007270/runat_04032004_160717_5098/app_query.fasta_1.3527
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10007270@cgn_1_1_37@runat_04032004_160717_5098 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PTCUS.COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 148.5 | 2.5 | 947 | 4 | US-09-418-780A-1 |
| 2 | 148.5 | 2.5 | 947 | 4 | US-09-392-714-23 |
| 3 | 146.5 | 2.5 | 2137 | 4 | US-09-134-001C-4463 |
| 4 | 146.5 | 2.5 | 2409 | 6 | 5180808-2 |
| 5 | 145.5 | 2.5 | 2035 | 2 | US-08-479-537A-2 |
| 6 | 145.5 | 2.5 | 2035 | 3 | US-09-083-116-2 |
| 7 | 145.5 | 2.5 | 2035 | 4 | US-09-134-916A-2 |
| 8 | 140.5 | 2.4 | 529 | 1 | US-08-178-477B-32 |
| 9 | 140.5 | 2.4 | 529 | 4 | US-09-304-121-2 |
| 10 | 140.5 | 2.4 | 783 | 4 | US-09-513-783A-176 |
| 11 | 139 | 2.4 | 1481 | 2 | US-08-616-844-40 |
| 12 | 139 | 2.4 | 1481 | 2 | US-08-599-654-40 |

| | | | | | | |
|----|-------|-----|------|---|----------------------|-------------------|
| 13 | 139 | 2.4 | 1481 | 3 | US-08-944-868A-40 | Sequence 40, Appl |
| 14 | 139 | 2.4 | 1481 | 3 | US-08-944-423A-40 | Sequence 40, Appl |
| 15 | 139 | 2.4 | 1481 | 3 | US-08-944-496-40 | Sequence 40, Appl |
| 16 | 136.5 | 2.3 | 985 | 5 | PCT-US96-03916-6 | Sequence 6, Appl |
| 17 | 136.5 | 2.3 | 985 | 5 | PCT-US96-03916-6 | Sequence 6, Appl |
| 18 | 135 | 2.3 | 1464 | 3 | US-08-891-640-2 | Sequence 2, Appl |
| 19 | 134.5 | 2.3 | 618 | 4 | US-09-252-991A-27666 | Sequence 27666, A |
| 20 | 134.5 | 2.3 | 1151 | 4 | US-09-328-352-4744 | Sequence 4744, Ap |
| 21 | 134 | 2.3 | 1087 | 4 | US-09-914-259-12 | Sequence 12, Appl |
| 22 | 132 | 2.2 | 2954 | 4 | US-09-150-867-1 | Sequence 1, Appl |
| 23 | 131 | 2.2 | 873 | 4 | US-09-540-824-28 | Sequence 28, Appl |
| 24 | 130.5 | 2.2 | 750 | 4 | US-09-540-236-3544 | Sequence 3544, Ap |
| 25 | 129 | 2.2 | 975 | 4 | US-09-914-259-19 | Sequence 19, Appl |
| 26 | 129 | 2.2 | 1420 | 2 | US-08-804-14 | Sequence 14, Appl |
| 27 | 129 | 2.2 | 1420 | 2 | US-08-218-265-14 | Sequence 14, Appl |
| 28 | 129 | 2.2 | 1420 | 3 | US-08-521-872-14 | Sequence 14, Appl |
| 29 | 129 | 2.2 | 1420 | 3 | US-08-590-399-14 | Sequence 14, Appl |
| 30 | 129 | 2.2 | 2500 | 3 | US-08-801-263A-2 | Sequence 2, Appl |
| 31 | 129 | 2.2 | 2500 | 3 | US-09-102-248-2 | Sequence 2, Appl |
| 32 | 129 | 2.2 | 2500 | 4 | US-09-367-764-2 | Sequence 2, Appl |
| 33 | 128.5 | 2.2 | 1494 | 3 | US-08-755-587-186 | Sequence 186, App |
| 34 | 128 | 2.2 | 1211 | 4 | US-09-134-001C-4820 | Sequence 4820, Ap |
| 35 | 128 | 2.2 | 3878 | 4 | US-09-914-259-11 | Sequence 11, Appl |
| 36 | 127.5 | 2.2 | 1190 | 4 | US-09-107-532A-7146 | Sequence 7146, Ap |
| 37 | 127.5 | 2.2 | 1202 | 1 | US-08-425-061-22 | Sequence 22, Appl |
| 38 | 127.5 | 2.2 | 1202 | 2 | US-08-825-886-22 | Sequence 22, Appl |
| 39 | 127.5 | 2.2 | 1202 | 4 | US-08-989-890-22 | Sequence 22, Appl |
| 40 | 127.5 | 2.2 | 1363 | 1 | US-08-425-061-23 | Sequence 23, Appl |
| 41 | 127.5 | 2.2 | 1363 | 2 | US-08-825-886-23 | Sequence 23, Appl |
| 42 | 127.5 | 2.2 | 1363 | 4 | US-08-989-890-23 | Sequence 23, Appl |
| 43 | 127.5 | 2.2 | 1852 | 1 | US-08-425-061-24 | Sequence 24, Appl |
| 44 | 127.5 | 2.2 | 1852 | 2 | US-08-825-886-24 | Sequence 24, Appl |
| 45 | 127.5 | 2.2 | 1852 | 4 | US-08-989-890-24 | Sequence 24, Appl |

ALIGNMENTS

RESULT 1
US-09-418-780A-1
; Sequence 1, Application US/09418780A
; Patent No. 6504009
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-043001
; CURRENT APPLICATION NUMBER: US/09/418,780A
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01782
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JAPAN 9/116402
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO.1
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-780A-1

Alignment Scores:
Pred. No.: 0.00322 Length: 947
Score: 148.50 Matches: 177
Percent Similarity: 33.44% Conservative: 124
Best Local Similarity: 19.67% Mismatches: 345
Query Match: 2.53% Indels: 254
DB: 4 Gaps: 37

US-10-007-270-1 (1-3330) x US-09-418-780A-1 (1-947)

Qy 179 CAAGTTCAGGACCAAGAT-----ATCTCCATTACATATCATCTTCT 223

Db 138 GlnValValGlyValGlyGluArgIleLysGlyThrGlnGlnAlaValSer 157

| | | | | |
|------|----|------|---|------|
| 224 | QY | 224 | GAACCTAAGACATAGACAATCCCCCAAGAACTGAAAGTACTGAAAAATG | 293 |
| | | | | |
| 158 | DB | 158 | SerAlaIysGluLysSerSerPro-----SerAlaThrGluLysVal | 171 |
| | | | | |
| 284 | QY | 284 | TACAAATGTCACATCATAGACGAGATATTCGATTGGCAAGCATCGAACAAAGATCC | 343 |
| | | | | |
| 172 | DB | 172 | PhelysGlnGlnGluIleProSerValPheProIysThr-SerIle----- | 186 |
| | | | | |
| 344 | QY | 344 | GCATTTTTCACCAACGGGGTTAAAGTGTCTCCACAGGAATCCATGAACAGATTTTAGAC | 403 |
| | | | | |
| 187 | DB | 187 | -----SerProLeuAsn----- | 190 |
| | | | | |
| 404 | QY | 404 | AGTCTTCAGCTTATATPAGATTGAGAGTGTGTGAGAAAGCAGATATGGGAAGCATATCGG | 463 |
| | | | | |
| 191 | DB | 191 | -----ValValGlnGlyAlaSerValAsnSerSe | 200 |
| | | | | |
| 464 | QY | 464 | ATCTTTTCGATCGCAGTCCCTCGACACAGGGGATATCAGG----- | 503 |
| | | | | |
| 200 | DB | 200 | xSerGlnThrAlaAlaGlnValThrLysGlyValLysArgLysAlaAspThrThrPr | 220 |
| | | | | |
| 504 | QY | 504 | -----ACTGGGTTCAGCATCTCCACAGCAGAGACCTTCTGCTCTTTGACATTCGAAAAAC | 559 |
| | | | | |
| 220 | DB | 220 | oAlaThrSerAlaValLysAlaSerSerGluPheSerProThrPheThr---GluLysSe | 239 |
| | | | | |
| 560 | QY | 560 | TTCCAGCAAT-----CCAGAGACACCTGGATCTTCTCCAGCAG | 598 |
| | | | | |
| 239 | DB | 239 | rValAlaLeuProProIleLysGluAsnMetProLysAsnValLeuProAspSerGlnG | 259 |
| | | | | |
| 599 | QY | 599 | AGATAAAAC-----AGAGAAGTTTCCCTGACAGAAAGATGAAATATTCGACAGAA- | 651 |
| | | | | |
| 259 | DB | 259 | ngInTyraenValValGluThrValLysValThrGluGlnLeuArgHis-CysSerGluI | 279 |
| | | | | |
| 652 | QY | 652 | -----GACATTTGGGAGAGCCTGG----- | 669 |
| | | | | |
| 279 | DB | 279 | leLeuLysGluMetLeuAlaLysLysHisPheSerTyraIatrpProPheTyraenProV | 299 |
| | | | | |
| 670 | QY | 670 | -----TGAACCATTTGCTCATTTCCACAGCAATCTACATTTCAAGACT | 712 |
| | | | | |
| 299 | DB | 299 | alaspValAsnAlaLeuGlyLeuHisAsnTyryrAspValValLysAsnProMetAspL | 319 |
| | | | | |
| 713 | QY | 713 | TGGCGCAGTATTCTAAGAAA----- | 732 |
| | | | | |
| 319 | DB | 319 | euGlyThrIleLysGluLysMetAspAsnGlnLysLysAspAlaTySerPheAlaA | 339 |
| | | | | |
| 733 | QY | 733 | -----CCCTCAGAGAG----- | 744 |
| | | | | |
| 339 | DB | 339 | laaspValArgLeuMetPheMetAsnCysTyryLysTyraenProProAspHisGluValV | 359 |
| | | | | |
| 745 | QY | 745 | -----CAAATTCAGATTTGCCACAGCTCTCACTTGGGCGCTTCCCTCTCA | 790 |
| | | | | |
| 359 | DB | 359 | alThrMetAlaArgMetLeuGlnAsnValPheGluThrHisPheSerLysIleProIleG | 379 |
| | | | | |
| 791 | QY | 791 | CTCCTGATGACAC-----CTCCTCAATGAAATTTCCGATTAATACACTCA | 835 |
| | | | | |
| 379 | DB | 379 | luProValGluSerMetProLeuCysTyryLleLysThrAspIleThrGluThrThrGlyA | 399 |
| | | | | |
| 836 | QY | 836 | ACGACACACAGATGCCTCAACAGAAAAGAAACAGAATTCGCTGTGTGGAGACGACA | 895 |
| | | | | |
| 399 | DB | 399 | rgGluAsnThrAsnGlnAlaSerSerGlnGluLysAsnSerSerAspAspSerGluAspGluA | 419 |
| | | | | |
| 896 | QY | 896 | GGGTGGAGCTCAGCTCTCTCTGTTAAACGAGATTCGAGGACGAGCTCGCTGACTCC | 955 |
| | | | | |
| 419 | DB | 419 | rgVal-----LysArgLeuAlaLysLeuGlnGlnL | 430 |
| | | | | |
| 956 | QY | 956 | AGTCCCCATATACCAGGAGCTAGCAGGAAGCTCCCACTTCAGATCGAAAGATATTTA | 1015 |
| | | | | |
| 430 | DB | 430 | eulysAlaValHisGlnGlnLeuGlnValLeuSerGlnValProPheArgLysLeuAsnL | 450 |
| | | | | |
| 1016 | QY | 1016 | AGAACTTCAGGATTCAAAAATTCATGTCTGTAGGATTTAGACCAAGAAAGAAAAAG | 1075 |
| | | | | |
| 450 | DB | 450 | ysLysLysGluLysSerLysLys-----GluLysLysGluLysV | 464 |
| | | | | |
| 1076 | QY | 1076 | ATGGCTCAAGCTCC-----ACAGAGATGCACTTACGGCCATCT | 1114 |

| | | | |
|----|------|--|------|
| Db | 464 | alaAenSerAenGluAenProAArgLysMetCysGluGlnMetArgLysGluLysGluLysS | 484 |
| QY | 1115 | TTAAGAGACACAGTGCAGAGCAAAAGCCCTGCAAGTGACCTCTGCTTTGATTCCA | 1174 |
| Db | 484 | erLysArgAen-----GlnProLysLysArgLysGlnGlnPheIleGly- | 498 |
| QY | 1175 | ACAAAATTGAAGTGAAGAGTCTATCATGACCAACCATGAGCGGACACAGCAACACAGAA | 1234 |
| Db | 499 | -----LeuLysSerGluAap-----GluApsAlaLysProMetA | 511 |
| QY | 1235 | TCTATCTCACAGCTACAGACCTCAAAAGCGTGATCAGCAAGACCTAGAGAGAAACAAT | 1294 |
| Db | 511 | enTyI-----AspGluLysArgGlnLeuSerLeuAsnIleAsnLysLeuProG | 527 |
| QY | 1295 | CTTTGGATGTGGGACAAATTCAGTTCACTGATCAAAATTCGTGGATCACTGCCAGCCTTTG | 1354 |
| Db | 527 | lyAspLysLeuGlyArgVal-----ValHisIleIleGlnSerArgLysProSerLeuS | 545 |
| QY | 1355 | GTCTTGACACCAATCAGAGCTGCCACATCTTTTGCTGTATACAGAGGATGCTACTT | 1414 |
| Db | 545 | erAenSerAenProAspGluIleGlnLysAspPheGluThrLeu---LysAlaSerThrL | 564 |
| QY | 1415 | TGAGTCAGCAACTTCTCTCTGTGNAACCCAGCTTGACAGCTGAGCGAGCAGAGCATG | 1474 |
| Db | 564 | euArg-----GluLeuGluLysTyValSerAla-----C | 574 |
| QY | 1475 | GTCTACTGACACTTCTTGCTCTCCACTCTCT-----ATGGCTCTCACTCCC | 1522 |
| Db | 574 | ysLeuArgLysArgProLeuLysProAlaLysLysIleMetMetSerLysGluGluL | 594 |
| QY | 1523 | TGTCAGAGCTCCACCTTCTTTTATGCAATCAAGCATCTCTCTGCACTGATCAAGGCA | 1582 |
| Db | 594 | euHiserGlnLysLysGlnLysGlnLysArgLeuLeuAspValAenAenGlnLeuA | 614 |
| QY | 1583 | CCACAGATACAATGGCCACTGACAGACAACTGATAGTACAGGGCTCACCATCCCC-- | 1639 |
| Db | 614 | enSerArgLysArgGlnThrLysSerAspLysThrGlnProSerLysAlaValGluAenV | 634 |
| QY | 1640 | CAAGTGATTAATTCGAATCAGCAACTGCTCTGGAAATTCATCATCCACTCTCATCTT | 1699 |
| Db | 634 | alSerArgLeuSerGluSerSerSerSerSerSerSerSerSerGluSerGluSerSerS | 654 |
| QY | 1700 | CAGATGAC-----AGCCGATCAAGTGCAGGTGGCGAAGATATGTCAGACACCTAG | 1750 |
| Db | 654 | erSerAspLeuSerSerSerAspSerSerAspSerGluSerGluMetPheProLysPheT | 674 |
| QY | 1751 | ATGAATATGGATCTGTCTACACTCTCTGCCCATCTGAGGTACCAGAGCTCAGCGAATATG | 1810 |
| Db | 674 | hrGluValLysProAenAspSerProSerLysGluHisValLysLysMetLysAenGluC | 694 |
| QY | 1811 | TTTCTGTCCAGAT-----CATTTCTGGAGATACCA | 1843 |
| Db | 694 | ysLeuLeuProGluLysArgThrGlyValThrGlnIleGlyTyCysValGlnAenPheTr | 714 |
| QY | 1844 | CTCTGTCTC---TCAGCTTACAGTATATCACCACCTAGTCTTATGACCATGCCGCCAAGG | 1900 |
| Db | 714 | hrSerAlaAenThrThrLeuValHisGlnThrThrProSerHisValMetProAenH | 734 |
| QY | 1901 | GCCAGAGCTGTGTAGTGTCTTCAGTCTCGTGTGTGAACATGGCCCTTCTCCACAGACC | 1960 |
| Db | 734 | iHisGlnLeuAla----- | 738 |
| QY | 1961 | TGTTTCAACAGAGCTCTCTGAGTACCAGCTCTGAGGACAACTACACA-----CAGC | 2014 |
| Db | 739 | ---PheAsnTyrGlnGluLeuGluHisLeuGlnThrValLysAsnIleSerProLeuGlnI | 758 |
| QY | 2015 | TGCTGGTTCCA-----TATCTAGATCCAATCTTACAGGATTI----- | 2052 |
| Db | 758 | leLeuProProSerGlyAspSerGluGlnLeuSerAenGlyIleThrValMetHisProS | 778 |
| QY | 2053 | -----AAGCAAC | 2059 |

```
Db 778 erGlyAspSerAspThrThrMetLeuGluSerGluCysGlnAlaProValGlnLysAspI 798
QY 2060 TTGAATACTTAACTTCAGAAACGGAGGTGTGATTGTGAATAGCAAAATGAGTTTGCTA 2119
Db 798 lElYslelYsAsn-----AlaAspSerTlPlySserLeuGlyL 811
QY 2120 AGTCGTGTC---CCGTATACCTCACCAGGCTGTGCACGGGGCTTCGGAGGATTTTCGTT 2176
Db 811 ysProValYsProSerGlyValMetLysSerSerAspGluLeuPheAsnGlnPheArgL 831
QY 2177 CTGCTGCA-----GCCCAACAACTC-----CATCTGG 2203
Db 831 ysAlaAlaIleGluLysGluValLysAlaArgThrGlnGluLeuIleArgLysHisLeuG 851
QY 2204 AAATAGACACTACTCTCAACATTGAACACGCTGATCAAGCAGATCCCTGCAAGTTCC 2263
Db 851 lUgInAsnThrLysGluLeuLysAlaSerGlnGluAsnGlnArgAspLeuGlyAsnGlyL 871
QY 2264 TGGCTCGCGCGAATTGCCCCAATGTGTAAGAACGAACGACTGAGGAGCGGAGTGTC 2323
Db 871 euThrValGluSerPheSerAsnLysIleGlnAsnLysCysSerGlyGluGluGln-Lys 890
QY 2324 GCTCCAAACACAGATATGACAGCGAGGAGCGCTGCGAGCTCTGGAACACGCGCTCTGTG 2383
Db 891 GluHisProGlnSerSerGluAlaGlnAspLysSerLysLeuIleuLysAspArg 910
QY 2384 GCCCTGCGCAAGCAATGCGAGTCTCTCCAGGAAAGGAGCTCCATGCAAGT 2437
Db 911 AspLeuAlaArgProLysGluGlnGluArgArgArgGluAlaMetValGly 928

RESULT 2
US-09-392-714-23
; Sequence 23, Application US/09392714A
; Patent No. 6686147
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Gure, Ali O.
; APPLICANT: Williamson, Barbara
; APPLICANT: Chen, Yao-Tseung
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Cancer Associated Antigens and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: L0461/7062
; CURRENT APPLICATION NUMBER: US/09/392,714A
; CURRENT FILING DATE: 1999-09-09
; EARLIER APPLICATION NUMBER: PCT/US98/14679
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-392-714-23
```

```
Alignment Scores:
Pred. No.: 0.000322 Length: 947
Score: 148.50 Matches: 177
Percent Similarity: 33.44% Conservative: 124
Best Local Similarity: 19.67% Mismatches: 345
Query Match: 2.53% Indels: 254
Db: 4 Gaps: 37
```

US-10-007-270-1 (1-3330) x US-09-392-714-23 (1-947)

```
QY 179 CAAGTTCAGGAACCAAGAT-----ATCTCCATTAAATATACCATCTCT 223
Db 138 GlnValValGlyValLysGluArgLleLysGlyThrGlnGlnAsnIleAlaValSer 157
QY 224 GAAACTAAAGACATAGACAATCCCCCAAGAAATGAACAATGAAAGTACTGAAAAATG 283
Db 158 SerAlaLysGluLysSerSerPro-----SerAlaThrGluLysVal 171
```

```
QY 284 TACAAATGTCACACTATGACGCAATATTGCAATTTGGCAAAAGCATCGACCAAAAGATCC 343
Db 172 PheLysGlnGlnGluIleProSerValPheProLysThr-Serlle----- 186
QY 344 GCATTTTCCCAACGGGGTTAAAGTCTGTCCACAGAAATCCATGAACACAGATTTTAGAC 403
Db 187 -----SerProLeuAsn----- 190
QY 404 AGTCTTCAAGCTTATTATAGATTGAGAGTGTGCAGAGCAGTATGGAGACATATCGG 463
Db 191 -----ValValGlnGlyAlaSerValAsnSerSe 200
QY 464 ATCTTCTTGGATCGCATCCCTGCACAGGGGAATATCAGG----- 503
Db 200 rSerGlnThrAlaAlaGlnValThrLysGlyValLysArgLysAlaAspThrThrPr 220
QY 504 ----ACTGGGTGACATCTCCGACGACGACCTTCGCTCTTTGACATTTGAAAAAAC 559
Db 220 olaThrSerAlaValLysAlaSerSerGluPheSerProThrPheThr---GluLysSe 239
QY 560 TTCAGCAATT-----CCAGGAGCACCTGGATCTTCTCCAGCAG 598
Db 239 rValAlaLeuProProIleLysGluAsnMetProLysAsnValLeuProAspSerGlnG 259
QY 599 AGAATAAAC-----AGAGAAGTTTCCCTGACAGAAAGATGAAATATCTGCAGAGAA- 651
Db 259 nGlnThrAsnValValGluThrValLysValThrGluGlnLeuArgHis-CysSerGluI 279
QY 652 -----GACATTGGGAGAGCCTCG----- 669
Db 279 leLeuLysGluMetLeuAlaLysLysHisPheSerTyrAlaTrpProPheTyrAsnPro 299
QY 670 -----TGAACACATTGTCAATTCAACAGCAATCTACATTTCAAGACT 712
Db 299 alAspValAsnAlaLeuGlyLeuHisAsnTyrTyrAspValValLysAsnProMetAspL 319
QY 713 TGGCGAGTATCTTAAGAAA----- 732
Db 319 euGlyThrIleLysGluLysMetAspAsnGlnGluTyrLysAspAlaTyrSerPheAla 339
QY 733 -----CCCTCAGAGAG----- 744
Db 339 laAspValArgLeuMetPheMetAsnCysTyrLysTyrAsnProAspHisGluVal 359
QY 745 -----CAAAATCAAGATGTGCCAACGTCCTCACTGGGCTTTCCCTCTCA 790
Db 359 alThrMetAlaArgMetLeuGlnAspValPheGluThrHisPheSerLysIleProIleG 379
QY 791 CTCTGTGATGACACC-----CTCCTCAATGAAATTCGTGAATATACACTCA 835
Db 379 luProValGluSerMetProLeuCysTyrIleLysThrAspIleThrGluThrThrGly 399
QY 836 AGACACCAAGATGCTTACACAGAAAGACAGAAATTCGCTGTGTGGAGGAGCAGA 895
Db 399 rgGluAsnThrAsnGluAlaSerSerGluGlyAsnSerSerAspAspSerSerGluAspGlu 419
QY 896 GGTGTGAGCTCAGCGTCTCTCTGTGTAAACACAGAAAGTTCAAGGACAGCTCGCTGACTCC 955
Db 419 rgVal-----LysArgLeuAlaLysLeuGlnGluL 430
QY 956 AGTCCCCATATTACAGGAGCTAGCAGAAAGTCCCACTTCAGATGCAAAAGATATTTA 1015
Db 430 euLysAlaValHisGlnGlnLeuGlnValLeuSerGlnValProPheAspGlyLeuAsnL 450
QY 1016 AGAACTTCCAGGATTCAAAAAATCCATGTGTGTAGGATTTAGACCACCAAGAAAGAAAG 1075
Db 450 ysLysLysGluLysSerLysLys-----GluLysLysGluLys 464
QY 1076 ATGGCTCAAGCTCC-----ACAGAGATGCAACTTACGGCCATCT 1114
Db 464 alAsnAsnSerAsnGluAsnProArgLysMetCysGluGlnMetArgLeuLysGluLysS 484
QY 1115 TTAAGACACACAGTGCAGACGCAAAAGCCCTGCAAGTGACCTCTGTCTTTTGTATCCA 1174
```

| | | | |
|----|------|---|------|
| Db | 484 | erLysArgAn-----GlnProLysLysArgLysGlnGlnPheIleGly----- | 498 |
| QY | 1175 | ACAAAATTCAAAGTCAGGAAGTCTATCATGCAACCATCGGAGGAGCAGCAGCAACGAA | 1234 |
| Db | 499 | -----LeuLysSerGluAsp-----GluAspAenAlaLysProMetA | 511 |
| QY | 1235 | TCATCTCACAGCTACAGACCTCAAAAGCGTATCAGCAAGCAAGCTACGAGCAAGACAAT | 1294 |
| Db | 511 | snTy-----AspGluLysArgGlnLeuSerLeuAenIleAsnLysLeuProG | 527 |
| QY | 1295 | CTTTGGAGTGGGCAACATTCAGTTCATGTGTAATTCGTGATCATCTGCCAGCCTTTG | 1354 |
| Db | 527 | lyAspLysLeuGlyArgVal-----ValHisIleIleGlnSerArgGluProSerLeuS | 545 |
| QY | 1355 | GTCTGTACACCAATCAGAGCTGCCACATCTTTTGTGTATTAACAGAGGATCTACTT | 1414 |
| Db | 545 | erAsnSerAenProAspGluIleGluLeuAspPheGluThrLeu---LysAlaSerThrL | 564 |
| QY | 1415 | TGAGTCCGAACATTCCTCTCTGTTGAACCCACTTGAGACAGTCGAGCAGCAGACGATG | 1474 |
| Db | 564 | euArg-----GluLeuGluLysTyValSerAla-----C | 574 |
| QY | 1475 | GTCTACCTGACACTCTTGGTCTCCACCTGCT-----ATGGCTCTTACTCTCC | 1522 |
| Db | 574 | ylLeuArgLysArgProLeuLysProAlaLysIleMetSerLysGluGluL | 594 |
| QY | 1523 | TGTCAGAAGCTCCACCTTTCTTATGTCATCAAGCATCTTCTCTGACTGATCAAGCA | 1582 |
| Db | 594 | euHisSerGlnLysGlnGluLeuGluLysArgLeuLeuAspValAsnAsnGlnLeuA | 614 |
| QY | 1583 | CACAGATACATGGCCACTGACGACAGCAATGCTAGPACGAGGCTCACATCCCTC--A | 1639 |
| Db | 614 | snSerArgLysArgGlnThrLysSerAspLysThrGlnProSerLysAlaValGluAsnV | 634 |
| QY | 1640 | CAAGTGATTATCTGCAATCAGCACTGGCTCTGGGAATTTTCATCTCACTGCTCATCTT | 1699 |
| Db | 634 | alSerArgLeuSerGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerS | 654 |
| QY | 1700 | CAGATGAC-----AGCCGATCAAGTGCAGGTGGCGAAGATATCGTCACAGCACCTAG | 1750 |
| Db | 654 | erSerAspLeuSerSerSerAspSerAspSerAspSerGluSerGluMetPheProLysPheT | 674 |
| QY | 1751 | ATGAATGGATCTGTCTGCATCTCTGCCCCATCTGAGGTACGAGCTCAGCAGTAATG | 1810 |
| Db | 674 | hrGluValLysProAsnAspSerProSerLysGluHisValLysLysMetLysAsnGluC | 694 |
| QY | 1811 | TTTCTGTCCCAGAT-----CATTTCTGGAGATACCA | 1843 |
| Db | 694 | ylIleLeuProGluGlyArgThrGlyValThrGlnIleGlyTySerValGlnAspThrT | 714 |
| QY | 1844 | CTCCTGTC--TCAGCTTTACGATATATCACCATAGTTCTATGACCATTCGCCCAAGS | 1900 |
| Db | 714 | hrSerAlaAsnThrThrLeuValHisGlnThrThrProSerHisValMetProProAsnH | 734 |
| QY | 1901 | GCCGAGAGCTGTGTAGTGTCTTCAGTCTGCGTGTCTTAACATGGCTCTCTCCAAGACC | 1960 |
| Db | 734 | isHisGlnLeuAla----- | 738 |
| QY | 1961 | TGTTCAACAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACA-----CAGC | 2014 |
| Db | 739 | --PheAsnTyGlnGluLeuGluHisLeuGlnThrValLysAenIleSerProLeuGlnI | 758 |
| QY | 2015 | TGCTGGTTCCA-----TATCTACGATCCCACTTACAGGATTT----- | 2052 |
| Db | 758 | IeLeuProProSerGlyAspSerGluGlnLeuSerAsnGlyIleThrValMetHisProS | 778 |
| QY | 2053 | -----AAGCAAC | 2059 |
| Db | 778 | erGlyAspSerAspThrThrMetLeuGluSerGluCysGlnAlaProValGlnLysAspI | 798 |
| QY | 2060 | TTGAATACTTAACTTCAGAAACGGGAGTGTGATGTGATAGCAAAATGAAGTTTGCTA | 2119 |

798 leLysIleLysAsn-----AlaAspSerTrpLysSerLeuGlyL 811

2120 AGCTGTGTG---CCGTATAACCTCACCAAGCTGTGCACGGGGTCTTGGAGGATTTTCGTT 2176

811 ysProValLysProSerGlyValMetLysSerAspGluLeuPheAsnGlnPheArg 831

2177 CTGTGTGCA-----GCCCAACAACCTC-----CATCTGG 2203

831 ysAlaAlaIleGluLysGluValLysAlaArgThrGlnGluLeuIleArgLysHisLeuG 851

2204 AAATGACAGCCTACTCTCTCAACATTGAACACAGCTGATCAAGCAGATCCCTGCACAGTCTC 2263

851 LuGlnAenThrLysGluLeuLysAlaSerGlnGluAsnGlnArgAspLeuGlyAsnGlyL 871

2264 TGGCTCTGGCGGAATTTCCCAATGTGTAAAGAACGAGACTGAGGAAGCGGAGTGTCT 2323

871 euThrValGluSerPheSerAsnLysIleGlnAsnLysCysSerGlyGluGluGln-Lys 890

2324 GCTCAAAACACGAGATATGACAGCCAGGGAGCCTGGACGCTCTGGACACAGGCTCTGTG 2383

891 GluHisProGlnSerSerGluAlaGlnAspLysSerLysLeuTrpLeuLeuLysAspArg 910

2384 CCCTGTGCACAAAGAAATGGAGGTCTCTCAGGAAGAGGAGCTCCATCGAGGT 2437

911 AspLeuAlaArgProLysGlyGluGlnGluArgArgArgGluAlaMetValGlyL 928

RESULT 3

US-09-134-001C-4463

; Sequence 4463, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4463

; LENGTH: 2137

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4463

| Alignment Scores: | |
|---|------------------|
| Pred. No.: | 0.000752 |
| Score: | 146.50 |
| Percent Similarity: | 35.23% |
| Best Local Similarity: | 20.81% |
| Query Match: | 2.49% |
| DB: | 4 |
| | Length: |
| | Matches: 2137 |
| | Matches: 124 |
| | Conservative: 86 |
| | Mismatches: 298 |
| | Indels: 89 |
| | Gaps: 15 |
| US-10-007-270-1 (1-3330) x US-09-134-001C-4463 (1-2137) | |

| | | | |
|----|------|--|------|
| QY | 221 | TCGGAACCTAAGACATAGACAATCCCCCAAGAAATCAAAACACTGAAGTACTGAAAA | 280 |
| | | | |
| Db | 1104 | SerAspSerAlaSerThrSerThrSerGluSerAspSerThrSerGluSerThrSerLeu | 1123 |
| | | | |
| QY | 281 | ATGTACAAATGTCACACTATGAGACGATATTCGATTGGCAAGCATCGCAACAAGA | 340 |
| | | | |
| Db | 1124 | SerGluSerLeuSerThr-----SerValSerAspSerThrSerAlaSerThrSerGlu | 1141 |
| | | | |
| QY | 341 | TCCGCATTTTCCCAACGGGGTTAA----- | 366 |
| | | | |
| Db | 1142 | SerAlaSerThrSerThrSerGluSerGluSerAsnSerAlaSerThrSerLeuSerGly | 1161 |
| | | | |
| QY | 367 | ACTCTGTCCACAGGAATCCATGAA-----ACAGATTTTAGACGTCTCAAGCTTA | 417 |
| | | | |
| Db | 1162 | SerLeuSerThrSerIleSerAspSerThrSerThrSerThrSerAspSerAlaSerThr | 1181 |
| | | | |

| | | | |
|--|------|---|------|
| QY | 1459 | GAGGAGCAGAGCATGGTCTACCTGACACACTTCTTGCTCTCCACCTGCTATGCGCTCTACC | 1511 |
| Db | 1524 | SerThrSerAspSerGluSerThrSerThrSerThrSerAspSerAlaSerThrSerThr | 1543 |
| QY | 1519 | TCCCTGTCTCAGAGCTCCACCTTCTTTATGGCATCAGCATCTCTCTCTGACTGATCAA | 1578 |
| Db | 1544 | SerValSerGlu | 1556 |
| QY | 1579 | GGCACCCAGATACAATGGCCCACTGACACAGACAATGCTAGTACCAGGGCTCACCATCCC | 1638 |
| Db | 1557 | GlySerThrSerThrSerValSerAspSerThr | 1570 |
| QY | 1639 | ACCATGTTATTCTGCAATCAGCCCACTGGCTCTGGGAATTCACATCCACTGCACTCT | 1698 |
| Db | 1571 | ThrSerAspSerAlaSerAlaSer | 1582 |
| QY | 1699 | TCAGATCAGACCGCATCAAGTCGAGGTGGCGAAGATATGGTCAGACACCTAGATGAATG | 1758 |
| Db | 1583 | AspSerAspSerAlaSerThrSerSerGluSerVal | 1597 |
| QY | 1759 | GANTGTCTGACACTCTCTGCCCCCATCTGAGGTACCGAGCTCAGCGAATATGTTCTGTC | 1818 |
| Db | 1598 | SerValSerAspSerThrSerAlaSerThrSerGluSerAlaSerThrSerThrSerVal | 1617 |
| QY | 1819 | CCAGATCAT | 1839 |
| Db | 1618 | SerAspSerAsnSerAlaSerThrSerSerLeuSerGluSerThrSerThrSerLeuSerAsp | 1637 |
| QY | 1840 | ACCATCTCTCTCAGCTTTACAGTATATACACCTAGTTCT | 1891 |
| Db | 1638 | SerThrSerMetSerThrSerAspSerAlaSerThrSerThr | 1651 |
| RESULT 4 | | | |
| 5180808-2 | | | |
| ; Patent NO. 5180808 | | | |
| ; APPLICANT: RUOSLAHTI, ERKKI I. | | | |
| ; TITLE OF INVENTION: VERISCAN CORE PROTEIN, NUCLEIC ACID | | | |
| ; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSION | | | |
| ; ANTIBODIES, AND METHODS OF DETECTING THE SAME | | | |
| ; NUMBER OF SEQUENCES: 4 | | | |
| ; CURRENT APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US/07/441,179 | | | |
| ; FILING DATE: 27-NOV-1989 | | | |
| ; SEQ ID NO:2: | | | |
| ; LENGTH: 2409 | | | |
| 5180808-2 | | | |
| Alignment Scores: | | | |
| Pred. No.: 0.008083 Length: 2409 | | | |
| Score: 146.50 Matches: 184 | | | |
| Percent Similarity: 29.66% Conservative: 115 | | | |
| Best Local Similarity: 18.25% Mismatches: 313 | | | |
| Query Match: 2.49% Indels: 396 | | | |
| DB: 6 Gaps: 48 | | | |
| US-10-007-270-1 (1-3330) x 5180808-2 (1-2409) | | | |
| QY | 691 | AGCAATCTACATTTCAAGACTTGGCGAGTATCTTAGAAAAACCTCAGAGAGCAAAATT | 750 |
| Db | 1468 | SerSerThrThrThrPheValSerAspGlySerLeuGluLysHisPro | 1484 |
| QY | 751 | CAAGATGTTGGCAACGCTCCTCCTCTGGCGCTTCCCTCTCACT | 792 |
| Db | 1485 | ProSerAlaLysAlaValThrAlaAspGlyPheProThrValSerValMetLeuProLeu | 1504 |
| QY | 793 | -----CCTGATGACACCTCTCTCAAT | 813 |
| Db | 1505 | HisSerGluGlnAsnLysSerSerProAspProThrSerThrLeuSerAsnThrValSer | 1524 |
| QY | 814 | -----GAAATCTCGAATAATACA | 831 |
| Db | 1525 | TyrGluArgSerThrAspGlySerPheGlnAspArgPheArgGluPheGluAspSerThr | 1544 |

Db 2183 AspTyrGlyTrpHisLysPheGlnGlyGlnCysTyrLysTyrPheAlaHisArgArgThr 2202
 QY 2440 CCAGATCCTCTGAAATCAA-----GCATACAAACTAGTGTAAAG 2484
 Db 2203 TrpAspAlaAlaGluArgGluCysArgLeuGlnGlyAlaHisLeuThrSerIleLeuSer 2222
 QY 2485 TTCCAAATCAACA-----AAT 2502
 Db 2223 HisGluGlnGlnMetPheValAsnArgValGlyHisAspTyrGlnTrpIleGlyLeuAsn 2242
 QY 2503 AACAGTAAATCAGTAA-----AGAAATCTGAATTAAGTACGCTAGATATGAGAA 2556
 Db 2243 AspLysMetPheGluHisAspPheArgTrpThrAspGlySerThrLeuGlnTyrGluAsn 2262
 QY 2557 TTT-----AACATCAAGAT----- 2571
 Db 2263 TrpArgProAsnGlnProAspSerPheSerAlaGlyGluAspCysValValIleIle 2282
 QY 2572 TGGGAAGGAATTAATAAAGTAA-----ATGTACAAATTAACCTTAGGCTAT-----CTC 2622
 Db 2283 TrpHisGluAsnGlyGlnTrpAsnAspValProCysAsnTyrHisLeuThrTyrThrCys 2302
 QY 2623 AGAGAGATGATTG-----CCTTCTCAAGAAATGGAGACAGGCATAT 2667
 Db 2303 LysLysGlyThrValAlaCysGlyGlnProProValValGluAsnAla-----LysTh 2320
 QY 2668 TCATGGCTCATCAAAATCCAGACATACAGTCAACACTGAGAAATCAGCACACACCATATTT 2727
 Db 2320 rPheGlyLysMetLysProArgTyrGluIleAsnSer----- 2332
 QY 2728 CAATATAGAGAGTATGATCTGTGCAACCACTAAATCTGAAAAAAGACACTTACT 2787
 Db 2333 -----LeuIleArgTyrHisCysLysAspGlyPhe-----IleGlnArgHisLeuPr 2348
 QY 2788 TATTATTAA----- 2802
 Db 2348 oThrIleArgCysLeuGlyAsnGlyArgTyrAlaIleProLysIleThrCysMetAsnPr 2368
 QY 2803 AAATGCATCAGCAACATAT 2824
 Db 2368 oSerAlaTyrGlnArgThrTyr 2375

RESULT 5
 US-08-479-537A-2
 ; Sequence 2, Application US/08479537A
 ; Patent No 5861381
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAMBER, Pierre
 ; APPLICANT: KIENY, Marie-Paule
 ; APPLICANT: LATHE, Richard
 ; APPLICANT: HAREUVENI, Mara
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 ; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 ; STREET: P.O. Box 1404
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/479,537A
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 90/13101
 ; FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1899
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA, or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
 OTHER INFORMATION: 21 amino acid precursor sequence."
 US-08-479-537A-2

Alignment Scores:
 Pred. No.: 0.00897 Length: 2035
 Score: 145.50 Matches: 74
 Percent Similarity: 38.11% Conservative: 51
 Best Local Similarity: 22.56% Mismatches: 144
 Query Match: 2.48% Indels: 59
 DB: 2 Gaps: 13

US-10-007-270-1 (1-3330) x US-08-479-537A-2 (1-2035)
 QY 1357 CTGACACCAATCAGAGTCCCATCTTTGCTGTATATACAGAGGATCTACTTTG 1416
 Db 1622 ProAsp***Arg---Pro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1640
 QY 1417 AGTCCAGAACTTCCTCT-----GTTGACCCAGCTTGAGACGCGA 1464
 Db 1641 AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1660

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 0.000897 | Length: | 2035 |
| Score: | 145.50 | Matches: | 74 |
| Percent Similarity: | 38.11% | Conservative: | 51 |
| Best Local Similarity: | 22.56% | Mismatches: | 144 |
| Query Match: | 2.48% | Indels: | 59 |
| DB: | 3 | Gaps: | 13 |

US-10-007-270-1 (1-3330) x US-09-083-116-2 (1-20335)

| | | | |
|----|------|---|------|
| QY | 1357 | CTGTGACACCACCAATCAGAGTCGCCACACTTTTGGTGCTGTATTACAGAGGATGCTACTTTG | 1411 |
| Db | 1622 | ProAsp****Arg-----Pro***ProGlySerThrAlaPro***AlaHisGlyValThrSer | 1640 |
| QY | 1417 | AGTCAGAGACTTCTCTCT-----GTTGAACCCCGAGCTTGAGACAAGTGACGCGA | 1464 |
| Db | 1641 | AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer | 1660 |
| QY | 1465 | GCAGAGCATGCTCTACCTGACACT---TCTTGGTGCTCCACTGCTATGGCTCTCACTCCC | 1521 |
| Db | 1661 | AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer | 1680 |
| QY | 1522 | CTGTCAAGAGCTCCACCTTCTTTATGGCATCAAGCACTTCTCTCTGACTGATCAAGGC | 1581 |
| Db | 1681 | AlaProAsp***ArgPro-----*****ProGly | 1689 |
| QY | 1582 | ACCACAGATACATGSCCACTGCACCAAGCAATGCTA-----GTACAGGG | 1626 |
| Db | 1690 | SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***ProGly | 1709 |
| QY | 1627 | CTCACCATCCCCACCAAGT-----GATTATTCTGCAATCAGCAACTGCTCTGGGA | 1677 |
| Db | 1710 | SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***LeuGly | 1729 |
| QY | 1678 | ATTTCACATCCACT-----GCATCTTCAGATGACAGCCGATCAAGTCAGAGTGGC | 1728 |
| Db | 1730 | SerThrAlaProProValHisAsnValThrSerAlaSerGlySerAlaSerGlySerAla | 1749 |
| QY | 1729 | GAAATATGTCAGACACTAGATGAATGGATCTCTCGACACTCTCTGCCCATCTTGAG | 1788 |
| Db | 1750 | SerThrLeuValHisAsnGlyThrSerAlaAlaThrThrProAlaSerIleSer | 1769 |
| QY | 1789 | GTACAGAGCTCAGCAATATGTTCTCCACAGATCATTTCTTGGAGGATACCACTCCT | 1848 |
| Db | 1770 | ThrProPro-----SerIleProSerHisHisSerAspThrProThrThr | 1784 |
| QY | 1849 | GTC-----TCAGCTTACAGTAT | 1866 |
| Db | 1785 | LeuAlaSerHisSerThrylsthyrAspAlaSerSerThrHisHisSerThrValProPro | 1804 |
| QY | 1867 | ATCACCACTAGTTCTATGACATTCGCCACCGAGGCGGAGAGCTGGTAGTCTTTC--- | 1923 |
| Db | 1805 | LeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyValSerPhePhePhe | 1824 |
| QY | 1924 | ---AGTCTGGTGTGTAACTAGTGCCTTCTCCAACGACCTGTTCACAGAAGAGCTCTCTG | 1980 |
| Db | 1825 | LeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuLeuAspProSerThrAsp | 1844 |
| QY | 1981 | GAGTACCGAGCTCTGGAGCAAATTCACACACTGCTGGTTCCTATCTACGATCCAAAT | 2040 |
| Db | 1845 | TyrrTyrglnGluLeuGlnArgAspIleSerGluMetPheLeuGlnAsnTyrllysGlnGly | 1864 |
| QY | 2041 | CTTACAGGATTTAAGCAACTTGAAATCTTAATCTTCAGAAACGGGAGTGTATGTGMAAT | 2100 |
| Db | 1865 | ----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGlnSerValValaGln | 1882 |
| QY | 2101 | AGCAATAAGATTGCTAAGTCTGTGCCGTATAACCTCACCAAGGCTGTGCACGGGCTC | 2160 |
| Db | 1883 | LeuThrLeuAlaPheArgGluGly-----ThrIleAsnValHisAspVal | 1897 |
| QY | 2161 | -----TTGGAGGATTTTGGTTCGTGCTGCAGGCCAACCACTCACTCGGAATAGAC | 2211 |
| Db | 1898 | GluThrGlnPheAsnGlnTyrlsthrThrGluAlaAlaSerArgTyrrAsnLeuThrIleSer | 1917 |

| | | | |
|----|------|--------------------------|------|
| QY | 2212 | AGCTACTCTCTCAACATGAACCA | 2235 |
| | | : : | |
| Db | 1918 | AspValSerValSerHisValPro | 1925 |

RESULT 7

US-09-134-916A-2
 ; Sequence 2, Application US/09134916A
 ; Patent No. 6328956
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAMRON, Pierre
 ; APPLICANT: KIENY, Marie-Paule
 ; APPLICANT: LATHE, Richard
 ; APPLICANT: HAREUVENI, Mara
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
 ; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 ; STREET: P.O. Box 1404
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/09/134,916A
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/479,537
 ; FILING DATE: 07-JUN-1995
 ; APPLICATION NUMBER: FR 90/13101
 ; FILING DATE: 23-OCT-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/FR91/00835
 ; FILING DATE: 23-OCT-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/039,320
 ; FILING DATE: 04-APR-1993
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/403,576
 ; FILING DATE: 14-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Teskin, Robin L.
 ; REGISTRATION NUMBER: 35,030
 ; REFERENCE/DOCKET NUMBER: 017753-025
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 836-6620
 ; TELEFAX: (703) 836-2021
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2035 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FEATURES:
 ; NAME/KEY: Peptide
 ; LOCATION: 134
 ; OTHER INFORMATION: /note= "Amino acid 134 is Xaa
 ; OTHER INFORMATION: Xaa Xaa which is the codon for Phe or Ala
 ; OTHER INFORMATION: 128 to 1899 constitute a repeated region where
 ; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The nu
 ; OTHER INFORMATION: repeats varies from 1 to 40."
 ; FEATURE:

```

; OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
; OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA
; OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
; OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA
; OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-134-916A-2

Alignment Scores:
Pred. No.: 0.000897 Length: 2035
Score: 145.50 Matches: 74
Percent Similarity: 38.11% Conservative: 51
Best Local Similarity: 22.56% Mismatches: 144
Query Match: 2.48% Indels: 59
DB: 4 Gaps: 13

US-10-007-270-1 (1-3330) x US-09-134-916A-2 (1-2035)
QY 1357 CTGACACCAATCAGAGCTGCCACATCTTTGCTGTTATACAGAGGATGCTACTTGG 1416
DB 1622 ProAsp***Arg---Pro**ProGlySerThrAlaPro***AlaHisGlyValThrSer 1640
QY 1417 AGTCAGAACTTCTCTCT-----GTTGAACCCAGCTTGAGAGCTGACGCGA 1464
DB 1641 AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1660
QY 1465 GCAGAGCATGCTTACTGACACT---TCTGTCTCCACCTGATGCGCTTACCTCC 1521
DB 1661 AlaProAsp***ArgPro***ProGlySerThrAlaPro***AlaHisGlyValThrSer 1680
QY 1522 CTGTCAGAGCTCCACCTTCTTTATGTCATCAAGCATCTTCTCTGATGATCAAGC 1581
DB 1681 AlaProAsp***ArgPro-----**ProGly 1689
QY 1582 ACACAGATACATGGCCCTGACACACAAATCTA-----GTACACGGG 1626
DB 1690 SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***ProGly 1709
QY 1627 CTCACCATCCCAACAGT-----GATATCTGCACATCAGCCAGCTGCTGGGA 1677
DB 1710 SerThrAlaPro***AlaHisGlyValThrSerAlaProAsp***ArgPro***LeuGly 1729
QY 1678 ATTTCATCCACCT-----GCATCTTCAGATGACAGCCAGCTCAGGTGGC 1728
DB 1730 SerThrAlaProValHisAsnValThrSerAlaSerGlySerAlaSerGlySerAla 1749
QY 1729 GAAGATATGTCAGACCTAGATGAATGGATCTGTCTGACATCTCTGCCCATCTGAG 1788
DB 1750 SerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrProAlaSerLysSer 1769
QY 1789 GTACACAGCTCAGCGAATATGTTCTGTCACGATCATTTCTGGAGGATACCATCTCT 1848
DB 1770 ThrProPro-----SerileProSerHisHisSerAspThrProThrThr 1784
QY 1849 GTC-----TCAGCTTACAGTAT 1866
DB 1785 LeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSerThrValProPro 1804
QY 1867 ATCACACATGTTCTATGACATTGCCGCCAGCGCCGAGAGCTGGTGTCTTC--- 1923
DB 1805 LeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyValSerPhePhe 1824

```

```

QY 1924 ---AGTCTCGGTGTGCTTAACATGCTTCCAAAGACCTGTTCAACAAGAGCTCTCTG 1980
DB 1825 LeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAspProSerThrAsp 1844
QY 1981 GAGTACGAGCTCTGGAGCAACAATTCACACAGCTGCTGTTCCATATCTACGATCCAAT 2040
DB 1845 TyrTyrGlnGluLeuGlnArgAspIleSerGluMetPheLeuGlnAsnTyrLysGlnGly 1864
QY 2041 CTTACAGGATTTAAGCAACTTGAATATTAACCTTCAGAAACGGAGCTGCTATGTGAAT 2100
DB 1865 -----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGluSerValValGln 1882
QY 2101 AGCAAAATGACACTTTGCTTAAGTCTGTGCGGTATTAACCTCACCAGGCTGTGCGGGTTC 2160
DB 1883 LeuThrLeuAlaPheArgGlnGly-----ThrIleAsnValHisAspVal 1897
QY 2161 -----TTGGAGGATTTTCTGCTGCTCAGCCCAACAACCTCATCTGGAAATAGAC 2211
DB 1898 GluThrGlnPheAsnGlnTyrLysThrGluAlaAlaSerArgTyrAsnLeuThrIleSer 1917
QY 2212 AGCTACTCTCTCAACATTGAACCA 2235
DB 1918 AspValSerValSerHisValPro 1925

RESULT 8
US-08-178-477B-32
; Sequence 32, Application US/08178477B
; Patent No. 5756343
; GENERAL INFORMATION:
; APPLICANT: WU, CARL; CLOS, JOACHIM;
; APPLICANT: WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
; TITLE OF INVENTION: CELL STRESS
; TITLE OF INVENTION: TRANSCRIPTIONAL FACTORS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/178,477B
; FILING DATE: 07-JAN-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/617,910
; FILING DATE: 26-NOV-1990
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4103US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-178-477B-32
Alignment Scores:

```



```

QY      1552  TCAGCATCTCTCTCTGACTGAT---CAAGGCACCACAGATACAATG-----1596
Db      343  AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPro 362
QY      1596  -----
Db      363  SerProProThrSerThrProGluLysCysLeuSerValAlaCysLeuAspLysAsn 382
QY      1597  -----GCCACTGAC-----CAGCAGTCGTA 1617
Db      383  GluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeu 402
QY      1617  -----
Db      403  SerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerVal 422
QY      1618  ---GTACAGCGGCTCACCATCCACCACGCTGATTATTCGCAATCAGC---CAACTGGCT 1671
Db      423  ThrValProMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnLeu 442
QY      1672  CTGGGAATTTTCATCCACCTGCACTCTTCAGATGATGACCGCATCAAGTGCA---GCTGGC 1728
Db      443  LeuSerProGlnProProArgProProGluAlaGluAsnSerSerProAspSerGly 462
QY      1729  GAAGATATGGTCAGACAC-----CTAGAT-----GAA 1755
Db      463  LysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuLeuAspProGlySerValAsp 482
QY      1756  ATGGATCTCTGACACTCTCTGCGCCCATCTGAGGTACACAGCTCAGCGAATATGTTTCT 1815
Db      483  ThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu---GlySerTyrPheSer 501
QY      1816  GTCCAGATCATTTCTTGAGGATACCACTCTCTCTCAGCTTTACAGTATATCACCCT 1875
Db      502  GluGlyAspGlyPheAlaGluAspProThr-----IleSerLeuLeuThrGly 517
QY      1876  AGTTCTATGACCATTTGCGGCGGCGAGAGCTGGTAGTG 1917
Db      518  SerGlu-----ProProLysAlaLysAspProThrVal 538

RESULT 9
US-09-304-121-2
; Sequence 2, Application US/09304121
; Patent No. 6342596
; GENERAL INFORMATION:
; APPLICANT: Voelmy, Richard
; TITLE OF INVENTION: MOLECULAR REGULATORY CIRCUITS TO ACHIEVE
; TITLE OF INVENTION: SUSTAINED ACTIVATION OF GENES OF INTEREST BY A SINGLE STR
; FILE REFERENCE: 870109.409
; CURRENT APPLICATION NUMBER: US/09/304.121
; CURRENT FILING DATE: 1999-05-03
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-304-121-2

Alignment Scores:
Pred. No.: 0.00118 Length: 529
Score: 140.50 Matches: 111
Percent Similarity: 34.24% Conservative: 65
Best Local Similarity: 21.60% Mismatches: 185
Query Match: 2.39% Indels: 153
DB: 4 Gaps: 26

US-10-007-270-1 (1-3330) x US-09-304-121-2 (1-529)

QY      700  CATTTCGTAAC-----CAGCTTGGCGAGTATCTTAAGAAACCCCTCAGAGACGCAATTCAA 753
Db      60  TyrPheLysHisAsnAsnMetAlaSerPheValArgGlnLeuAsnMetTyrGlyPheArg 79

```


| | | | |
|----|------|---|------|
| QY | 901 | GAGCTCAGCGTCTCTCTGTGTAACACCAAGTATTCAGAGCAGAGCTCGCTGACTCCGAC--- | 957 |
| DB | 393 | LyseLeuLeuThrAspValGlnLeuMetIlyseGlyIleGlnGluCysMetAspSerIlyseLeu | 412 |
| QY | 958 | -----TCCCATATTACAGAGACTAGCAGGAAAGTCCCAA--- | 993 |
| DB | 413 | LeuAlaMetIlyseHisGluAsnGluAlaLeuTrpArgGluValAlaSerLeuArgGlnIlyse | 432 |
| QY | 994 | ---CTTCAGATGCAAAAGATATTTAAGAACTCCAGGATTC-----AAA | 1035 |
| DB | 433 | HisAlaGlnGlnIlyseValValAsnIlyseIleGlnPheLeuIleSerLeuValGln | 452 |
| QY | 1036 | AAATCCATGTTAGATTTAGACAAAG-----AAAGAAAAGATGCCTCA | 1083 |
| DB | 453 | SerAsnArgIleLeuGlyValIlyseArgIlyseProLeuMetLeuAsnAspSerGlySer | 472 |
| QY | 1084 | AGCTCCAGAGATGCACTTACGCGCATCTTTAGACACACACTGCAGAGACAAAAGC | 1143 |
| DB | 473 | AlaHisSerMetProIlyseTyrSerArgGlnPheSerLeuGluHisValHisGlySerGly | 492 |
| QY | 1144 | CCTGCAAGTACGCTCTGCTTTTGGATTCACCAAAATTTGAAGCTGAGGAAGTCTATCAT | 1203 |
| DB | 493 | Pro-----TyrSerAlaProSerProAlaTyrSerSerSerSerLeuTyr--- | 507 |
| QY | 1204 | GGAACATGAGGAGGACCAAGCAACCAAGAAATCTATCTCAGCAGCTACAGACCTCAAAGG | 1263 |
| DB | 508 | -----AlaProAspAlaValAlaSerSerGlyProIle----- | 518 |
| QY | 1264 | CTGATCAGCAACCATAGAGAGAGAACATCTTTGGATGTGGGGACAATTCAGTTCACT | 1323 |
| DB | 519 | -----IleSerAspIle | 522 |
| QY | 1324 | GATGAAATTCGTGGATCACTGCCAGCTTTTGGTCTCT---GACACCAATCAGAGCTGCC | 1380 |
| DB | 523 | ThrGluLeuAlaProAlaSerProMetAlaSerProGlyIlyseIleAspIlyseArgPro | 542 |
| QY | 1381 | ACATCTTTTCTGTATTAACAGAGAGATGCTACTTTGAGTCCAGAACTCTCTCTGTTGAA | 1440 |
| DB | 543 | LeuSerSerProLeuValArg-----ValIlyseGluGluProProSerPro | 558 |
| QY | 1441 | CCCCAGCTTCAGACAGTGGAGCGAGAGCATGTCTACTCT---GACACTTCT | 1491 |
| DB | 559 | ProGlnSerProArgValGluGluAlaSerProGlyIlyseIleAspIlyseArgPro | 578 |
| QY | 1492 | TGGTCTCCACTGCTGGCTCTTACCTCCCTGTCAGAGCTCCACCTTTCTTATGGCA | 1551 |
| DB | 579 | LeuSerProThrAlaLeuIleAspSerIleLeuArgGluSerGluPro-----AlaPro | 596 |
| QY | 1552 | TCAGCATCTTCTCTCACTGAT---CAAGCACCACAGATACAATG----- | 1596 |
| DB | 597 | AlaSerValThrAlaLeuThrAspAlaArgGlyHisThrAspThrGluGlyArgProPro | 616 |
| QY | 1596 | ----- | 1596 |
| DB | 617 | SerProProThrSerThrProGluIlyseCysLeuSerValAlaCysLeuAspLysAsn | 636 |
| QY | 1597 | -----GCCACTGAC-----CAGACAAATGCTA | 1617 |
| DB | 637 | GluLeuSerAspHisLeuAspAlaMetAspSerAsnLeuAspAsnLeuGlnThrMetLeu | 656 |
| QY | 1617 | ----- | 1617 |
| DB | 657 | SerSerHisGlyPheSerValAspThrSerAlaLeuLeuAspLeuPheSerProSerVal | 676 |
| QY | 1618 | ---GTACCGAGGCTCACATCCCACTGATTTCTGCAATCAGC---CAACTGGCT | 1671 |
| DB | 677 | ThrValProAspMetSerLeuProAspLeuAspSerSerLeuAlaSerIleGlnGluLeu | 696 |
| QY | 1672 | CTGGGAATTCATCCACTGTCATCTTCAGATGCACGCCGATCAAGTCA---GTGGC | 1728 |
| DB | 697 | LeuSerProGlnGluProProArgProProGluAlaGluAsnSerProAspSerGly | 716 |

```

QY 1729 GAAGATATGGTCACAC-----CTAGAT-----GAA 1755
Db 717 LysGlnLeuValHisTyrThrAlaGlnProLeuPheLeuAspProGlySerValAsp 736
QY 1756 ATGGATCTGTCTGACACTCTCCGCCCATCTGTAGGTACACGAGCTCAGCGAATATCTTCT 1815
Db 737 ThrGlySerAsnAspLeuProValLeuPheGluLeuGlyGlu---GlySerTyrPheSer 755
QY 1816 GTCCAGATCATTTCTTGTGAGGATACACATCTCTCTGTCTCAGCTTTACAGTATATACCACT 1875
Db 756 GlucGlyAspGlyPheAlaGluAspProThr-----IleSerLeuLeuThrGly 771
QY 1876 AGTTCTATGCACCATTCCTCCCAAGCGCGGAGAGCTGTAGTG 1917
Db 772 SerGlu-----ProProlsAlaLysAspProThrVal 782

RESULT 11
US-08-616-844-40
; Sequence 40, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,844
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 03-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-616-844-40

Alignment Scores:
Pred. No.: 0.00281 Length: 1481
Score: 139.00 Matches: 180
Percent Similarity: 32.48% Conservative: 113
Best Local Similarity: 19.96% Mismatches: 383
Query Match: 2.37% Indels: 226

```

DB: 2 Gaps: 39

US-10-007-270-1 (1-3330) x US-08-616-844-40 (1-1481)

QY 242 AATCCCCAAGAATAAGAAACAACCTGAA---AGTACTGAAAAAATGTACAAAATGTCAACT 298
Db |||||
QY 390 AsnProGlyAspGluGluPheLeuSerThrValSer 409
Db |||||
QY 299 ATGAGCAGGAATTTCGAT-----TTGGCAAAGCATCGAAACAAAAGATCCGCACATT 349
Db :|||
QY 410 LeuArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSerSerSerGlu 429
Db |||||
QY 350 TTCCCACACGGGGTAAAGTCTGTCCACAGAATCCATGAAACAGATTTTAGACAGTCTT 409
Db |||||
QY 430 ValGlnAsnGlySerProMetSerGlnThrGluThrVal-SerArg----- 444
QY 410 CAAGCTTATTAGATTGAGTGTTGACGAGACAGTATGGGAAGCATATCGGATCTTT 469
Db |||||
QY 445 -----SerValAlaProMetArgGlyGlyGluIleThrAlaHi 457
QY 470 CTGATCGCA-----TCCCTGACACAGGGGAATATCAGGACTGGGTACAGATCTGCCAG 523
Db |||||
QY 457 sTpLeuLeuThrAsnSerThrThrSerAlaAspValThrGlySerSerAlaSerTyPr 477
QY 524 CAGGAGACCTTCGCTCT--TTGACATTTGGAAAAAATTTCAGCAATTTCCOAG-- 575
Db |||||
QY 477 oGluGlyValAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyd 497
QY 576 -AGCACCTGGATCTTCCACAGAGAAATAAACACAGAGAAGTTTCCCTGCACAGAAAAGAT 634
Db |||||
QY 497 ySerHisThrAlaLeuGlyAspArgSerySerGluSerSerSerThrSerSerSer 516
QY 635 GAAATATCTGCAGAGAGACATTTGGAGACCTGTGGTAACCATTCTCATTTCAACAGCA 694
Db |||||
QY 517 -----GluSerLeuAsnSerSerAlaProArgGlyGluAr 528
QY 695 ATCTACATTTCAAGACTTGGCGCATATTCATAGAAAACCTCAGAGAGCAAAATCAAG 754
Db |||||
QY 528 gSerThrLeuGlu-Asp-----SerArgGluProGlyGlnAlaLeuGlyAsps 544
QY 755 ATCTGCCAACGTC-----TCACTGGCGCTTCCTCTCA 790
Db |||||
QY 544 erSerAlaAsnAlaGluAspArgThrSerGlyValProSerLeuGlyThrHisThrLeuA 564
QY 791 CTCCT-----GATGACACCCCTCTCAATGAAATTCGATAATACATCAACAG 838
Db |||||
QY 564 laThrValThrGlyAsnGlyGluArgThrLeuArgSerVal-----ThrLeuThrA 581
QY 839 ACACCAAGATCCTACACAGAA-----AGAG 865
Db |||||
QY 581 snThrSerMetSerThrThrSerGlyGluAlaGlySerProAlaAlaAlaMetProGlnG 601
QY 866 AAACAGAAATTCGCTGTGTGGAGGAGCAGAGGGTGGAGCTCAGCTCTCTCTGGTAACC 925
Db |||||
QY 601 luThrGluGlyAlaSerLeuHis----- 608
QY 926 AGAGTTCAAGCAGAGCTTCGCTGCATCCOAGTCCCATATTACCAGGAGCTACGACGAA 985
Db |||||
QY 609 -----ValAsnValThrAspAspMetGlyLeuValSerArgSerLeuAlaAlaA 625
QY 986 AGTCCCACTTCAGATGCMAAAGATATTATAGAAACTTCAGGATTCAAAANAATCCATG 1045
Db |||||
QY 625 erSerAlaLeuGlyValAlaGlyIle-----SeryGlyGlnValArgSg 640
QY 1046 TGTTAGGATTTAGACCAAGAAAGAAAAGATGGCTCAAGCTCCACAGAGATGCACATTA 1105
Db |||||
QY 640 lyThrAlaIleGluGlnArgThrSerSerAspHisThrAspHisThr-----TyLeus 658
QY 1106 CGGCATCTTTTAAGACAC-----AGTGCAGAAAGCAAAAAGCGCTGCAG 1150
Db |||||
QY 658 erSerThrPheThrLysGlyGluArgAlaLeuSerIleThrAspAsnSerSerSerS 678
QY 1151 GTGACTCTGCTCTTTTGATTCCACAAAATTTGAAGGTAGGAGAGTCTATCATGTGGAACCA 1210

| | | |
|------|---|------|
| 1014 | eu1leProLeuThrSerValProThr-SerAlaLysGluMetThrThrLysLeu----- | 1031 |
| 2147 | CTGTCAACGGGCTCTGGAGGATTTTCGTTCTGTCGACGCCAACATCTCATCTGGA | 2206 |
| 1032 | -----GlyValThrAlaGluTyr-SerProAla----- | 1040 |
| 2207 | TAGACAGCTACTCTTCAACATTGAACACAGCTGATCAA----- | 2244 |
| 1041 | -----SerArgSerLeuGlyThrSerProSerProGlnThrThrValSerThrAlaG | 1059 |
| 2244 | ----- | 2244 |
| 1059 | luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrT | 1079 |
| 2245 | -----GCAGATCCCTCCAGTTCCTGCTCGCGCGAAATTGGCC | 2284 |
| 1079 | hrLeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCysLeuHisAsnGlyG | 1099 |
| 2285 | AATGCTAAAGAACGACGACTGAGGAGCGGAGTCTCGCTGCAACACGAGATATGACA | 2344 |
| 1099 | luCysValAlaAspAsnThrSerArgGlyTyrHisCysArgCysProSerTrp---- | 1117 |
| 2345 | GCCAGGGAGCGCTCGACGGTCTGCGAACCGACGCGCTCTGT-----GCCCTCGCA | 2392 |
| 1118 | --GlnGlyAspAsnCysSerSerValAspValAsnGluCysLeuSerAsnProCysProSerT | 1137 |
| 2393 | CAAGGAATGCGAGTCTCTCCAGGGAAGGAGCTCCATCGACGGTTGCCA----- | 2442 |
| 1137 | hrAlaThrCysAsnAsnThrGlnGlySerPheIle---CysLysCysProValGlyTyr-G | 1156 |
| 2443 | -----GATCATCTGAAATCAAGCATACAAACTAGTGTATAAAAGTTCCAAAAATCAAC | 2497 |
| 1156 | lnLeuGluLysGlyIleCysAsnLeuValArgThrPheValThrGluPhelLysLeuLysA | 1176 |
| 2498 | AAAAAT-----AACAGGTAATCAGTAAAGAAATTCGAATTACTGACCGTAGAATATG | 2551 |
| 1176 | rgThrPheLeuAsnThrThrValGluLys---HisSerAspLeuGlnValGluAsnG | 1195 |
| 2552 | AA 2553 | |
| 1195 | lu 1195 | |

RESULT 12
US-08-599-654-40
; Sequence 40, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: FALS, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITILE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSER: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,654
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995

```

; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-599-654-40

Alignment Scores:
Pred. No.: 0.00281 Length: 1481
Score: 139.00 Matches: 180
Percent Similarity: 32.48% Conservative: 113
Best local Similarity: 19.96% Mismatches: 333
Query Match: 2.37% Indels: 26
DB: 2 Gaps: 39

US-10-007-270-1 (1-3330) x US-08-599-654-40 (1-1481)

QY 242 AATCCCCCAAGAAATGAACAACTCGAA--AGTACTGAAAAAATGTACAAATGTCAACT 298
    |||||
Db 390 AsnProGlyAspGluGluPheIleGluProSerThrGluAsnGluPheGlyLeuThrSer 409

QY 299 ATGAGACAGAAATTCGAT-----TTGCCAAAGCATCGAACAAAAAGATCCGCATT 349
    |||||
Db 410 LeuArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSerSerSerGlu 429

QY 350 TTCCCAACGGGGITAAAGTCTGCTCCACAGAAATCCATGAACACAGATTTTAGACAGTCT 409
    |||||
Db 430 ValGlnAsnGlySerProMetSerGlnThrGluThrVal-SerArg-----444

QY 410 CAGCTTTATTATAGATTGAGATGTGTGTCAGGAAGCAGTATGGGAGCATATCGGATCTTT 469
    |||||
Db 445 -----SerValAlaProMetArgGlyGlyGluIleThrAlaHi 457

QY 470 CTGGATCGCA-----TCCTGCACACAGGGGAATATCAGGACTGGTCAGCATCTGCCAG 523
    |||||
Db 457 sTrpLeuLeuThrAsnSerThrThrSerAlaAspValThrGlySerSerAlaSerTyrrPr 477

QY 524 CAGGAGACCTTCCTCCTCT--TTGACNTTGGAAAAAAGCTTCAGAAATCCCAAG----- 575
    |||||
Db 477 oGluGlyValAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyGly 497

QY 576 -AGCACTTGGATCTTCTCCACAGAGAAATAAACACAGAGAGATTTCCTGTGACAGAAAGAT 634
    |||||
Db 497 ySerHisThrAlaLeuGlyAspApgSerTyrrSerGluSerSerSerThrSerSer--516

QY 635 GAAATATCTGCAGACAGACATTCGGGAGAGCCTGGTGAACCATTTGTCATTTCAACAGCA 694
    |||||
Db 517 -----GluSerLeuAsnSerSerAlaProArgGlyGluArg 528

QY 695 ATCTACATTCAAAGACTTGGCGCAGTATTCTTAAGAAAAACCTCAGAGACAGCAAAATCAAG 754
    |||||
Db 528 gSerThrLeuGlu-Asp-----SerArgGluProGlyGlnAlaLeuGlyAsps 544

QY 755 ATGTTGCCAAGTC-----TCACTTGGGCCCTTTCCCTCTCA 790
    |||||
Db 544 exSerAlaAsnAlaGluAspArgThrSerGlyValProSerLeuGlyThrHisThrLeuA 564

QY 791 CTCCT-----GATGACACCTCTCAATGAATTTCTCGATAATACACTCAACG 838
    |||||
Db 564 laThrValThrGlyAsnGlyGluArgThrLeuArgSerVal-----ThrLeuThrA 581

QY 839 ACACCAAGATGCTACACAGAA-----AGAG 865
    |||||

```

Db 581 snThrSerMetSerThrThrSerglyGluAlaGlySerProAlaAlaMetProGlnG 601
 QY 866 AAACAGAAATCGCTGTGTTGGAGGAGCAGAGGCTGAGCTCAGCGTCTCTCTGGTAACC 925
 Db 601 luthrGluGlyAlaSerLeuHis 608
 QY 926 AGAAGTTCAAGGAGAGCTCGCTGACTCCAGTCCCAATATACAGAGCTAGCAGGAA 985
 Db 609 -----ValAenValThrAspAspMetGlyLeuValSerArgSerLeuAlaLas 625
 QY 986 AGTCCCAACTTCAGATGCMAAAGATATTTAAGAAATCTCCAGGATTCMAAAATATCCATG 1045
 Db 625 erSerAlaLeuGlyValAlaGlyLe 640
 QY 1046 TGTAGATTAGACCAAGAAAGAAAGATGGCTCAAGCTCCACAGAGATGCAACTTA 1105
 Db 640 lylThrAlaIleGluGlnArgThrSerSerAspHisThr 658
 QY 1106 CGGCCATCTTTAGAGACAC-----AGTGCAAGCAAAAGCCCTGCAA 1150
 Db 658 erSerThrPheThrLysGlyGluAlaLeuSerIleThrAspAenSerSerS 678
 QY 1151 GTGACCTCTCTGCTTTTGTATCCAAACAAATTTGAAGTGAGGAGTCTATCATGGAACCA 1210
 Db 678 erAspIleValGluSerSerThrSerTyrlleLysIleSerAenSerSerGluT 698
 QY 1211 TGGAGGAGCAGCAACCAAGAAATCTATCTCACAGCTCACAGCTCAAAAGGCTGATCA 1270
 Db 698 yrSerSerPheSerHisAlaGln-----ThrGluArgSerAenIleSerS 713
 QY 1271 GCAAGCACTAGAGGAGCAAACTCTTGGATGTGGGCAATTCAGTCTCACTGATGAA 1330
 Db 713 ertyrAspGlyGlyTyrlAlaGlnProSerThrGluSerProValLeuHisThrSerAsnL 733
 QY 1331 TTGCTGGATCATCTGCCAGCCTTTGCT---CCTGACACCAATCAGAGCTGCCACATCTT 1387
 Db 733 euProSerTyrlThrProThrIleAenMetProAenThrSerValValLeuAspThrAspA 753
 QY 1388 TTGCTCTTATACAGAGGATGCTACTTTGAGT----- 1419
 Db 753 laGluPheValAspSerSerSerSerSerSerSerSerSerSerS 773
 QY 1420 --CCAGAACTTCTCTCTGTTGAACCCAGCTTGAGACATGAGCGAGCAGAGCAT- 1473
 Db 773 lyProProLeuPro-----LeuProSerValSerGlnSerHisLeuLeuP 788
 QY 1474 -----GGTCTACCTGACACTTCTTGGTCTCCACTGATGCGCTCTACCTCCCTGT 1525
 Db 788 heSerSerIleLeuProSerThrArgAlaSerValHisLeuLeuLysSerThrSerAspA 808
 QY 1526 CAGAAGCTCCACCTTCTTTATGGCATCAGACATCTTCTCTGACTGATCAAGGCACCA 1585
 Db 808 laSerThrPro-----TrpSerSerSerProSerProLeuProValSerLeuThr 825
 QY 1586 CAGATACAAATGGCACTGACAGCAATGCTAGTACAGGGCTCACCTCCACCCACCATG 1645
 Db 825 hrSerThrSerAla-----ProLeuSerValSerGlnThrLeuProGlnSerS 842
 QY 1646 ATATCTGCAATCAGCCAACTGGCTCTGGGAATTTTCATCCACTGCTGATCTTCAGATG 1705
 Db 842 erSerThrProValLeuProArgAla-----ArgGluThrProValThrSerPheG 859
 QY 1706 ACAGCGGATCAGTGCAGGTGGCAGATATGTCAGACCTAGATGAATGATCTGT 1765
 Db 859 lnThrSerThrMetThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuL 879
 QY 1766 CTGACACTCTCTGCCCATCTGAGGTACAGAGCTCAGCGAA-----TATG 1810
 Db 879 ysSerGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProSerLeuV 899
 QY 1811 TTTCTGCCAGATCAATCTTCTGGAGGATCACCTCTCTGCTCAGCTTTCAGTATATCA 1870
 Db 899 alSerLeuProThrGluSerThrLysAlaValThrThrAsnSerProLeuProProSerL 919

QY 1871 CACTAGTCTTATGACCATGCCCCCAAGGCGGAGAGCTGTAGTGTCTTCTCAGTCTGC 1930
 Db 919 euThrGluSerSerThr-----GluGlnThrLeuProAlaThrSerThrA 934
 QY 1931 GTGTTGCTAAACATGGCC-----TTCCTCAACGACCTGTTCACACAGAGC----- 1974
 Db 934 snLeuAlaGlnMetSerProThrPheThrThrThrIleLeuLysThrSerGlnProLeuM 954
 QY 1975 -----TCTCTGAGTACCGAGCTCTGGGCGAAC 2002
 Db 954 etThrThrProGlyThrLeuSerSerThrAlaSerLeuValThrGlyProIleAlaValG 974
 QY 2003 AATTCACA-----CAGCTGCTGGTTCCATATCTA- 2031
 Db 974 llnThrThrAlaGlyLysGlnLeuSerLeuThrHisProGluIleLeuValProGlnIleS 994
 QY 2032 -----CGATCCAATCTTACAGATTTAAGCACTTGAATACTTAACCTTCAGAAACGGGA 2086
 Db 994 erThrGluGlyGlyIleSerThrGluArgAsnArgValIleValAlaGlnAlaThrThrGlyL 1014
 QY 2087 GTGTGATGTGAATAGCAAAATGAAGTTTCTTAAGTCTGTGCGGTATACCTTCACCAAGG 2146
 Db 1014 euIleProLeuThrSerValProThrSerAlaLysGluMetThrThrLysLeu----- 1031
 QY 2147 CTGTGCACGGGCTCTGAGGATTTCTGCTGTGCGAGCCCAACCACTCCATCTCGAAA 2206
 Db 1032 -----GlyValThrAlaGluTyrlSerProAla----- 1040
 QY 2207 TAGACACTACTCTCTCAACATTCGAACAGCTGATCAA----- 2244
 Db 1041 -----SerArgSerLeuGlyThrSerProSerProGlnThrThrValValSerThrAlaG 1059
 QY 2244 ----- 2244
 Db 1059 luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrT 1079
 QY 2245 -----GCAGATCCCTGCAAGTTCTCTGGCTGGCGGCAATTTGCC 2284
 Db 1079 hrLeuSerSerSerAlaSerValAsnSerCysAlaValAenProCysLeuHisAenGlyG 1099
 QY 2285 AATGTGTAAGAAACCAACGAGCTGAGAGCGGAGTGTGCTGCAACACAGGATATGACA 2344
 Db 1099 luCysValAlaAspAsnThrSerArgGlyTyrlHisCysArgCysProSerTrp----- 1117
 QY 2345 GCAGAGGAGCTCGAGCGTCTGGAACACAGGCTCTGT-----GGCCTTGGA 2392
 Db 1118 --GlnGlyAspAspCysSerValAspValAenGluCysLeuSerAsnProCysProSerT 1137
 QY 2393 CAAAGGAATGCGAGTCTCTCCAGGAAAGGAGGCTCCATGCGAGGTGGCA----- 2442
 Db 1137 hrAlaThrCysAsnAenThrGlnGlySerPheIle---CysLysCysProValGlyTyrlG 1156
 QY 2443 -----GATCCTCTGAAAATCAAGCATACAAAATAGTGTAAAGTTCCAAAATCAAC 2497
 Db 1156 lnLeuGluLysGlyLeCysAenLeuValArgThrPheValThrGluPheLysLeuLysA 1176
 QY 2498 AAAAT-----AACAGGTAATCAGTAAAGAAATTTCTGAATTTACTGACCTAGATATG 2551
 Db 1176 rgThrPheLeuAenThrThrValGluLys---HisSerAspLeuGlnGluValGluAenG 1195
 QY 2552 AA 2553
 Db 1195 lu 1195

RESULT 13

US-08-944-868A-40
 ; Sequence 40, Application US/08944868A
 ; Patent No. 6018025

GENERAL INFORMATION:

; APPLICANT: FALB, DEAN A
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE

/ NUMBER OF SEQUENCES: 54
 / CORRESPONDENCE ADDRESS:
 / ADDRESS: PENNIE & EDMONDS
 / STREET: 1155 Avenue of the Americas
 / CITY: New York
 / STATE: New York
 / COUNTRY: USA
 / ZIP: 10036-2711
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent In Release #1.0, Version #1.30
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/944,868A
 / FILING DATE:
 / CLASSIFICATION:
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 08/599,654
 / FILING DATE:
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 08/386,844
 / FILING DATE: 10-FEB-1995
 / ATTORNEY/AGENT INFORMATION:
 / NAME: CORUZZI, LAURA A.
 / REGISTRATION NUMBER: 30,742
 / REFERENCE/DOCKET NUMBER: 7853-041
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (212) 790-9090
 / TELEFAX: (212) 869-8864
 / TELEX: 66141 PENNIE
 / INFORMATION FOR SEQ ID NO: 40:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 1481 amino acids
 / TYPE: amino acid
 / STRANDEDNESS:
 / TOPOLOGY: unknown
 / MOLECULE TYPE: protein
 / US-08-944-868A-40

Alignment Scores:
 Pred. No.: 0.00281 Length: 1481
 Score: 139.00 Matches: 180
 Percent Similarity: 32.48% Conservative: 113
 Best Local Similarity: 19.96% Mismatches: 383
 Query Match: 2.37% Indels: 226
 DB: 3 Gaps: 39

US-10-007-270-1 (1-3330) x US-08-944-868A-40 (1-1481)

QY 242 AATCCCAAGAAATGAAACTGAA---AGTACTGAAATAATGTACAAATGTCAACT 298
 Db |||||
 QY 390 AnProGlyAspGluGluPheIleGluProSerThrGluAsnGluPheGlyLeuThrSer 409
 Db |||||
 QY 299 ATGAGACGAATATTCGAT-----TTGGCAAGACATCGAACAAAGATCCGCAATT 349
 Db |||||
 QY 410 LeuArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSerSerGlu 429
 Db |||||
 QY 350 TTCCCAACGGGGTTAAAGTGTGTCCACAGGAATCATGAAACAGATTTTAGACAGCTTT 409
 Db |||||
 QY 430 ValGlnAsnGlySerProMetSerGlnThrGluThrVal-SerArg----- 444
 QY 410 CAAGCTATTATAGATTGAGAGTGTCTCAGGAGCAGATGTGGGAACATATGGATCTTT 469
 Db |||||
 QY 445 -----SerValAlaProMetArgGlyGlyGluIleThrAlaHi 457
 QY 470 CTGGATGCGCA-----TCCCTGACACAGGGGAATATCAGGACTGGGTGAGCATCTGCCAG 523
 Db |||||
 QY 457 stPLeuLeuThrAsnSerThrSerAlaAspValThrGlySerSerAlaSerTyPr 477
 QY 524 CAGGAGACCTTCCTGCTCT---TTGACATTGGAAAAAACTTCAGCAATCCAGG----- 575
 Db |||||
 QY 477 oGluGlyValAlaAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyGI 497

QY 576 -AGCACCTGGATCTTCTCCAGCAGAGAAATAAAACAGAGAGTTCCTCCAGCAGAAAGAT 634
 Db |||||
 QY 497 ySerHisThrAlaLeuGlyAspArgSerTySerGluSerSerSerThrSerSerSer-- 516
 Db |||||
 QY 635 GAAATATCTGCAGAGAGACATTCGGAGAGCTGTGGTGAACCATTTGTCAATTTCAACAGCA 694
 Db |||||
 QY 517 -----GluSerLeuAsnSerSerAlaProArgGlyGluAr 528
 QY 695 ATCTACATTTCAAGACTTGGGACAGTATTTCTAAGAAAAACCTCAGAGAGGAGCAATTCAG 754
 Db |||||
 QY 528 gSerThrLeuGlu-Asp-----SerArgGluProGlyGlnAlaLeuGlyAspS 544
 QY 755 ATGTTGCCAAGTTC-----TCACTGGGCTTTTCCCTCTCA 790
 Db |||||
 QY 544 erSerAlaAlaAlaGluAspArgThrSerGlyValProSerLeuGlyThrHisThrLeuA 564
 QY 791 CTCCT-----GATGACACCTCTCAATGAATTCCTCGATATACACTCAACG 838
 Db |||||
 QY 564 laThrValThrGlyAsnGlyGluArgThrLeuArgSerVal-----ThrLeuThrA 581
 QY 839 ACACCAAGATGCTTACAAACAGAA-----AGAG 865
 Db |||||
 QY 581 enThrSerMetSerThrThrSerGlyGluAlaGlySerProAlaAlaAlaMetProGing 601
 QY 866 AACACAGAAATTCCTGTGTGGAGGAGCAGAGGTTGGAGCTCAGCGTCTCTCTGTGTAACC 925
 Db |||||
 QY 601 luthrGluGlyAlaSerLeuHis-----SeryrGlyGlnValArgG 640
 QY 926 AGAAGTTCAAGGCAGAGCTCGTCACTCCAGTCCCATATTTACACAGAGCTACAGGAA 985
 Db |||||
 QY 609 -----ValAsnValThrAspAspMetGlyLeuValSerArgSerLeuAlaAlas 625
 QY 986 AGTCCCACTTCAGATGCAAAAGATATTTAAGAACTTCAGGATTTCAAAAAATTCATG 1045
 Db |||||
 QY 625 erSerAlaLeuGlyValAlaGlyLe-----SeryrGlyGlnValArgG 640
 QY 1046 TGTTAGATTAGACCAACAGAAAGAAAGATGCTCAGCTCCACAGTCCACAGAGATGCACTTA 1105
 Db |||||
 QY 640 lyThrAlaIleGluGlnArgThrSerSerAspHisThrAspHisThr-----TyrLeuS 658
 QY 1106 CGGCCATCTTTAAAGAGACAC-----AGTGCAGAACCAAAAGCCCTGCA 1150
 Db |||||
 QY 658 erSerThrPheThrLysGlyGluArgAlaLeuLeuSerIleThrAspAsnSerSerSers 678
 QY 1151 GTGACCTCTCTCTTTTGTATCCAAATTTGAAAGTGAAGAGTCTATCATCGGAACCA 1210
 Db |||||
 QY 678 erAspIleValGluSerSerThrSerTyrlleLysIleSerAsnSerSerHisSerGluT 698
 QY 1211 TGGAGGAGGACACAGCAACAGAAATCTATCTCACAGCTACAGACCTCAAAAGCTGATCA 1270
 Db |||||
 QY 698 yrSerSerPheSerHisAlaGln-----ThrGluArgSerAsnIleSers 713
 QY 1271 GCAAGACACTAGAGAGAAACAATCTTTGGATGTGGGAGCAATTCAGTTCACCTGATGAA 1330
 Db |||||
 QY 713 erTyraSpGlyGluTyraAlaGlnProSerThrGluSerProValLeuHisThrSerAsnL 733
 QY 1331 TTGCTGGATCACTGCCAGCTTTTGTGT---CCTGACACCCCAATCAGAGTCCCAATCTT 1387
 Db |||||
 QY 733 euProSerTyThrProThrIleAsnMetProAsnThrSerValValLeuAspThrAspA 753
 QY 1388 TTGCTGTTATGACAGAGGATGCTACTTTGAGT----- 1419
 Db |||||
 QY 753 laGluPheValSerAspSerSerSerSerSerSerSerSerSerSerSerSerg 773
 QY 1420 --CCAGAACTTCCTCTCTGTGTGAACCCAGCTTGACAGTGGAGCGGAGCAGAGCAT--- 1473
 Db |||||
 QY 773 lyProProLeuPro-----LeuProSerValSerGlnSerHisLeuP 788
 QY 1474 -----GCTTACCTGACACTTCTTGTGTCTCCACCTGTGTATGGCTCTTACCTCCCTGT 1525
 Db |||||
 QY 788 heSerSerIleLeuProSerThrArgAlaSerValHisLeuLeuLeuSerSerSerAspA 808

QY 1526 CAGAAGCTCCACCTTCTTTATGGCATCAAGCATCTCTCTGACTGATCAAGGCACCA 1585
 Db 808 laSerThrPro-----TrrSerSerProSerProLeuProValSerLeuThrt 825
 QY 1586 CAGATACAAATGGCACTGACACAGATGCTAGTACAGGGCTCACCATCCCAACGATG 1645
 Db 825 hrSerThrSerAla-----ProLeuSerValSerGlnThrLeuProGlnSerS 842
 QY 1646 ATTATTCGCAATCAGCAACTGCTCTGGGAATTTCAATCCATCCCTGCTGATGATG 1705
 Db 842 erSerThrProValLeuProArgala-----ArgGluThrProValThrSerPheG 859
 QY 1706 ACAGCCCATCAAGTGGAGTGGCGAAGATATGTCAGACACTAGATGAATGATCTGT 1765
 Db 859 lnThrSerThrMetThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuL 879
 QY 1766 CTGACACTCTCGCCCATCTGAGGTACAGACTCAGCGAA-----TATG 1810
 Db 879 ysSerGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProSerLeuV 899
 QY 1811 TTTCTGTCCTCAGATCATTTCTTGAGGATACACCTCTGCTCAGCTTTTACAGTATATCA 1870
 Db 899 alSerLeuProThrGluSerThrLysAlaValThrThrAsnSerProLeuProProSerL 919
 QY 1871 COACTAGTCTATGACCATTCGCCCAAGCGCGAGAGCTGGTAGTGTCTTCAGTCTGC 1930
 Db 919 eutThrGluSerSerThr-----GluGlnThrLeuProAlaThrSerThrA 934
 QY 1931 GTGTGTCTACATGGCC-----TTCTCCCAACGACTGTTCAACAGAGC----- 1974
 Db 934 enLeuAlaGlnMetSerProThrPheThrThrThrIleLeuLysThrSerGlnProLeuM 954
 QY 1975 -----TCTCTGGAGTACCGAGCTCTGGAGCAAC 2002
 Db 954 etThrThrProGlyThrLeuSerSerThrAlaSerLeuValThrGlyProIleAlaValG 974
 QY 2003 AATTCACA-----CAGCTGCTGTTCCATCTA- 2031
 Db 974 lnThrAlaGlyLysGlnLeuSerLeuThrHisProGluIleLeuValProGlnIleS 994
 QY 2032 -----CGATCCCAATTTACAGGATTTAAGCACTGGAATTAATTAATTCAGAAACGGA 2086
 Db 994 erThrGluGlyLysLeuThrGluArgAsnArgValIleValAspAlaThrGlyL 1014
 QY 2087 GTGTGATTTGTAATGCAAAATGAGTTTGTCTAGTCTGCGGTATTAATTAATTCACCAAG 2146
 Db 1014 eulleProLeuThrSerValProThrSerAlaLysGluMetThrThrLysLeu----- 1031
 QY 2147 CTGTGCACGGGTCTTGAGGATTTTCTGCTGCTGCGCCCAACCACTCCATCTCGAA 2206
 Db 1032 -----GlyValThrAlaGluThrSerProAla----- 1040
 QY 2207 TAGACAGCTACTCTCACTGATGACAGCTGATCAA----- 2244
 Db 1041 -----SerArgSerLeuGlyThrSerProSerProGlnThrThrValValSerThrAlaG 1059
 QY 2244 ----- 2244
 Db 1059 luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrT 1079
 QY 2245 -----GCAGATCCCTCAAGTTCTCTGCGCTGGCGGGAATTTGCC 2284
 Db 1079 hrLeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCysLeuHisGlyG 1099
 QY 2285 AATGTCTTAAGNACGACGAGCTGAGGAGCGAGTGTCTGCGAAACCCAGGATATGACA 2344
 Db 1099 luCysValAlaAspAsnThrSerArgGlyTyriHisCysArgCysProProSerTrp---- 1117
 QY 2345 GCCAGGGAGGCTGACGCTCTGGAACCCAGGCTCTGT-----GGCCTTGCCA 2392
 Db 1118 --GlnGlyAspAspCysSerValAspValAlaGluCysLeuSerAsnProCysProSerT 1137
 QY 2393 CAAGGAATGCGAGTCTCTCCAGGGAAGGAGCTCCATGCGAGTGTCCCA----- 2442

Db 1137 hrAlaThrCysAsnAsnThrGlnGlySerPheIle-----CysLysCysProValGlyTyrg 1156
 QY 2443 -----GATCACTCTGAAATCAAGCATACAACTAGTGTATAAAGTTCCAAATCAAC 2497
 Db 1156 lnLeuLysGlyLysCysAsnLeuValArgThrPheValThrGluPheLysLeuLysA 1176
 QY 2498 ABAAT-----AACAGGTATTCAGTAAAGAAATTTCTGAATTAATGACCGTAGAATATG 2551
 Db 1176 rgThrPheLeuAsnThrThrValGluLys---HisSerAspLeuGlnGluValGluAsnG 1195
 QY 2552 AA 2553
 Db 1195 lu 1195
 RESULT 14
 US-08-944-423A-40
 ; Sequence 40, Application US/08944423A
 ; Patent No. 6020463
 ; GENERAL INFORMATION:
 ; APPLICANT: FALB, DEAN A
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
 ; NUMBER OF SEQUENCES: 54
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: PENNIE & EDMONDS
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: WINDOWS 95
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/944,423A
 ; FILING DATE: 06-OCT-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/599,654
 ; FILING DATE: 09-FEB-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/485,573
 ; FILING DATE: JUN-07-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/386,844
 ; FILING DATE: 10-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CORUZZI, LAURA A
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7853-105
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1481 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-944-423A-40
 Alignment Scores:
 Pred. No.: 0.00281
 Score: 139.00
 Length: 1481
 Percent Similarity: 32.48%
 Matches: 180
 Best Local Similarity: 19.96%
 Mismatches: 113
 Query Match: 2.37%
 Indels: 226
 Gaps: 39

| | | | |
|------|---|---|------|
| 242 | ATATCCCCCAAGAAATGAACAACTGAA | ---AGTACTGAAAAAATGTACAAATGCTCACT | 298 |
| Db | | | |
| 390 | AsnProGlyAspGluGluPheIleGluProSerThrGluAsnGluPheGlyLeuThrSer | 409 | |
| QY | | | |
| 299 | ATCAGACAGAAATATTCGAT | ---TTGGCAAGAGCATCGAACAAAAAAGATCCGCAATTT | 349 |
| Db | | | |
| 410 | LeuArgTyrGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSerSerSerGlu | 429 | |
| QY | | | |
| 350 | TTCCCAACGGGGTAAAGTCTGTCCACAGGAATCCATGAACAGATTTAGACAGTCTT | 409 | |
| Db | | | |
| 430 | ValGlnAsnGlySerProMetSerClnThrGluThrVal | ---SerArg | 444 |
| QY | | | |
| 410 | CAAGCTTATTATAGATTGAGAGTGTCGACGAAGCAGATGGAAGCATATCGGATCTTT | 469 | |
| Db | | | |
| 445 | -----SerValAlaProMetArgGlyGlyIleThrAlaHis | 457 | |
| QY | | | |
| 470 | CTGGATCCGA | ---TCCCTGACACAGGGGAATATCAGACTGGGTGAGCATCTCCAG | 523 |
| Db | | | |
| 457 | StrpLeuLeuThrAsnSerThrThrSerAlaAspValThrGlySerSerAlaSerTyrP | 477 | |
| QY | | | |
| 524 | CAGGAGACCTTCGCTCT | ---TTGACATTCGMAAAACTTCAGCAATCCGAG | 575 |
| Db | | | |
| 477 | OgluGlyValAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyI | 497 | |
| QY | | | |
| 576 | -AGCACCTGGATCTTCCACGACAGAGATAAACAGAGAAGTTTCCCTGCACAAAAAGAT | 634 | |
| Db | | | |
| 497 | YSerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerThrSerSerSer | 516 | |
| QY | | | |
| 635 | GAATATCTGACAGAGACATTCGACAGAGCTGGTCAACCACTTGTTCATTTCACACAGA | 694 | |
| Db | | | |
| 517 | -----GluSerLeuAsnSerSerAlaProArgGlyGluArg | 528 | |
| QY | | | |
| 695 | ATCTACATTTCAAAGACTTTGGCAGTATCTTAAAGAAACCTTCAGACAGACAAATTCAG | 754 | |
| Db | | | |
| 528 | GserThrLeuGluAsp | -----SerArgGluProGlyGlnAlaLeuGlyAsp | 544 |
| QY | | | |
| 755 | ATGTGTCCAACGTC | -----TCACCTGGGCCTTTCCCTCTCA | 790 |
| Db | | | |
| 544 | erSerAlaAsnAlaClnAspArgThrSerGlyValProSerLeuGlyThrHisThrLeuA | 564 | |
| QY | | | |
| 791 | CTCCT | -----GATGACACCTCTCCATGAATTTCTGATATACACTCAACG | 838 |
| Db | | | |
| 564 | lathrValThrGlyAsnGlyGluArgThrLeuArgSerVal | -----ThrLeuThrA | 581 |
| QY | | | |
| 839 | ACACCAAGATGCTTACACAGAA | -----AGAG | 865 |
| Db | | | |
| 581 | snThrSerMetSerThrThrSerGlyGluAlaGlySerProAlaAlaAlaMetProGlnG | 601 | |
| QY | | | |
| 866 | AAACAGAAATTCGCTGTGTGGAGGAGCAGAGGGTGAGCTCAGGCTCTCTGTGTAAAC | 925 | |
| Db | | | |
| 601 | luthrGluGlyAlaSerLeuHis | ----- | 608 |
| QY | | | |
| 926 | AGAAGTTCAGCGCAGAGCTCGCTGACTCCGCTCCAGTCCCGCCCATATTACACGAGCTAGCAGAA | 985 | |
| b | | | |
| 609 | -----ValAsnValThrAspAspMetGlyLeuValSerArgSerLeuAlaAla | 625 | |
| QY | | | |
| 986 | AGTCCCACTTCAGTGCAGAAAGATATTTAGAAATTCAGGATTTCAAAAATCCCATG | 1045 | |
| b | | | |
| 625 | erSerAlaLeuGlyValAlaGlyIle | -----SerTyrGlyGlnValArg | 640 |
| QY | | | |
| 1046 | TGTTAGGATTTAGACAAAGAAAGAAAGATGGCTCAAGCTCCACAGAGATGCACATTA | 1105 | |
| b | | | |
| 640 | lyThrAlaIleGluGlnArgThrSerSerAspHisThrAspHisThr | -----TyrLeuS | 658 |
| QY | | | |
| 1106 | CGGCATCTTTTAGACACAC | -----AGTCAGACAGCAAAAAGCCCTGCAG | 1150 |
| b | | | |
| 658 | erSerThrPheThrTyrGlyGluArgAlaLeuLeuSerIleThrAspAsnSerSerSerS | 678 | |
| QY | | | |
| 1151 | GTGACCTCTCTTTGATTTCCCAACAAAATTTGAAGTGGAGGAGCTCTATCTATCGGAACA | 1210 | |
| b | | | |

Df

678 exAsplleValGluSerSerThrSerTyrlleLyslleSerIasnSerSerHisSerGluT 698
QY 1211 TGGAGGAGGACAGCAACCAAGAATCTATCTCACAGCTCAGACACTCAAAGGCCTATCA 1270
Df 698 yrSerSerPheSerHisaladln-----ThrGluArgSerAsnIleSerS 713
QY 1271 GCAAGACACTTAGAGNAGAACAACTTTTGATGTGGGACATTAGTCCTACTGATGA 1330
Df 713 erTyraSpGlyGlnTyAladlnProSerThrGluSerProValLeuHisThrSerA 733
QY 1331 TTGCTGGATCATCGCCAGCCTTTGGF---CCTGACACCACAATCAGAGCTGCCACATCTT 1387
Df 733 euProSerTyThrProThrIleAsnMetProAnThrSerValValLeuAspThrAsp 753
QY 1388 TTGCTGTATAACAGAGATGCTACTTTGAGT----- 1419
Df 753 laGluPheValserAspSerSerSerSerSerSerSerSerSerSerSerS 773
QY 1420 --COAGAACTCTCTCTGTTTGAACCCAGCTTCAGACAGCTGGACGAGCAGAGCAT-- 1473
Df 773 lyProLeuPro-----LeuProSerValSerGlnSerHisHisLeuP 788
QY 1474 -----GGTCTACCTGACACTCTTGGTCTCCACTGCTATGGCTCTACTCCCTCT 1525
Df 788 heSerSerIleLeuProSerThrArgAlaSerValHisLeuLeuLysSerThrSerAsp 808
QY 1526 CAGAAGCTCCACCTTTCTTTATGGCANTCAGCATCTTCTCTGACTGATCAAGGCACCA 1585
Df 808 laSerThrPro-----TrpSerSerSerProSerProLeuProValSerLeuThrT 825
QY 1586 CAGATACAAATGGCCACTGACAGCAAACTGATGATACCAGGGCTCACATCCCACCACTG 1645
Df 825 hrSerThrSerAla-----ProLeuSerValSerGlnThrThrLeuProGlnSerS 842
QY 1646 ATTATCTGCAATCAGCCAATGCTGCTGGAAATTCACATCCACTGCATCTTCAGATG 1705
Df 842 eSerThrProValLeuProArgAla-----ArgGluThrProValThrSerPheG 859
QY 1706 ACAGCCGATCAAGTGCAGGTGGCGAAGATATGTCACAGACACTAGATGAATGATCTGT 1765
Df 859 lnThrSerThrMetThrSerPheMetThrMetLeuHisSerSerGlnThrAlaSpl 879
QY 1766 CTGACACTCTCCGCCCATCTGAGTACACAGACTCAGCGAA-----TATG 1810
Df 879 ysSerGlnSerThrProHisGlnGluLysValleThrGluSerLysSerProSerLeuV 899
QY 1811 TTCTGTGCCAGATCATTTCTTGAGAGATACCATCTCTGTCAGCTTTTACGATATATCA 1870
Df 899 alSerLeuProThrGluSerThrLysAlaValThrThrAsnSerProLeuProProSerL 919
QY 1871 COACTAGTCTTATGACCAATTCGCCCAAGGCGAGAGCTGCTAGTGTCTTCTCAGTCTGC 1930
Df 919 euThrGluSerSerThr-----GluGlnThrLeuProAlaThrSerThrA 934
QY 1931 GTGTGTCTTAACATGGCC-----TTCTCCACAGCACTGTTCAACAAGAC----- 1974
Df 934 snLeuAlaGlnMetSerProThrPheThrThrThrileLeuLysThrSerGlnProLeuM 954
QY 1975 -----TCTCTGAGTACCGAGCTCTGGAGCAAC 2002
Df 954 etThrThrProGlyThrLeuSerSerThrAlaSerLeuValThrGlyProPolelaValG 974
QY 2003 AATTCACA-----CAGCTGCTGGTTCCATATCTA- 2031
Df 974 lnThrThrAlaGlyLysGlnLeuSerLeuThrHisproGluIleLeuValProGlnIles 994
QY 2032 ----CGATCCAACTCTACAGATTAAAGCACTTGAANAATTAATCTTCAGAAACCGGA 2086
Df 994 erThrGlUGlyGlyIleSerThrGluArgAsnArgValleValAspAlaThrThrGlyL 1014
QY 2087 GTGTGATTGTGAATACAAAATGAAGTTTCTTAAGTCTGTGCGGTATACCTACCAAGG 2146
Df 1014 euIleProLeuThrSerValProThrSerAlaLysGluMetThrThrylsLeu----- 1031

QY 2147 CTGTGACAGGGCTTGGAGGATTTGGTTCTGCTGCGAGCCCAACAACTCCATCTGGAAA 2206
Db 1032 -----GlyValThrAlaGluTyr-SerProAla----- 1040
QY 2207 TAGACGCTACTCTCTCAACATTGAACGAGCTGATCAA----- 2244
Db 1041 -----SerArgSerLeuGlyThr-SerProSerProGlnThrThrValValSerThrAlaG 1059
QY 2244 ----- 2244
Db 1059 luAspLeuAlaProLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrT 1079
QY 2245 -----GCAGATCCCTGCAAGTTCTCTGGCTGCGGCGCAATTTGGCC 2284
Db 1079 hrLeuSerSerAlaSerValAsnSerCysAlaValAsnProCysLeuHisAsnGlyG 1099
QY 2285 AATGTGTGAAGAACGAGCTGAGGAGCGGAGTCTGCTGCAACAGATATGACA 2344
Db 1099 luCysValAlaAspAsnThrSerArgGlyTyrHisCysArgCysProSerTrp----- 1117
QY 2345 GCCAGGGAGCGCTGACGGCTGGAACACAGGCTCTGT-----GGCCTGGCA 2392
Db 1118 --GlnGlyAspAspCysSerValAspValAsnGluCysLeuSerAsnProCysProSerT 1137
QY 2393 CAAGGAATGCGAGTCTCCAGGGAAGGAGCTCCATGCGAGTTGCCA----- 2442
Db 1137 hrAlaThrCysAsnAsnThrGlnGlySerPheIle---CysLysCysProValGlyTyrG 1156
QY 2443 ----GATCACTCTGAAATCAAGCATACAACTAGTGTAAAGTTCCAAATCAAC 2497
Db 1156 lnLeuGluLysGlyIleCysAsnLeuValArgThrPheValThrGluPheLysLeuLysA 1176
QY 2498 AAAAT-----AACAGGTAATCAGTAAAGAAATCTGAATTAATGACCTAGATATG 2551
Db 1176 rgThrPheLeuAsnThrThrValGluLys---HisSerAspLeuGlnGluValGluAsnG 1195
QY 2552 AA 2553
Db 1195 lu 1195

RESULT 15

US-08-944-496-40
; Sequence 40, Application US/08944496
; Patent No. 6124433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,496
; FILING DATE: 06-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-944-496-40
Alignment Scores:
Pred. No.: 0.00281 Length: 1481
Score: 139.00 Matches: 180
Percent Similarity: 32.48% Conservative: 113
Best Local Similarity: 19.96% Mismatches: 383
Query Match: 2.37% Indels: 226
DB: 3 Gaps: 39
US-10-007-270-1 (1-3330) x US-08-944-496-40 (1-1481)
QY 242 AATCCCCCAAGAAATGAACAACTGAA---AGTACTGAAAAAATGTACAAATGTCAACT 298
Db 330 AsnProGlyAspGluGluPheIleGluProSerThrGluAsnGluPheGlyLeuThrSer 409
QY 299 ATGACGCGAATATTGCAT-----TTGCCAAGCATCGAACAAGAATCGCATTT 349
Db 410 LeuArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSerSerSerGlu 429
QY 350 TTCCCAACGCGGGTTAAAGTCTGTCCACAGGAATCCATGAACACAGATTTTAGACAGTCT 409
Db 430 ValGlnAsnGlySerProMetSerGlnThrGluThrVal-SerArg----- 444
QY 410 CAAGCTTATTATGAGATTGAGAGTGTGTGAGGAGCAGTATGGGAAAGCATATCGGATCTTT 469
Db 445 -----SerValalaProMetArgGlyGlyGluIleThrAlaHi 457
QY 470 CTGGATCGCA-----TCCCTGACACAGGGGAATATCAGACTGGGTCAGCATCTGCCAG 523
Db 457 sTrpLeuLeuThrAsnSerThrThrSerAlaAspValThrGlySerSerAlaSerTyrPr 477
QY 524 CAGGAGACCTTCTGCCTCT---TTGACATTGGAAAAAACTTCACCAATTTCCAGG----- 575
Db 477 oGluGlyValAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497
QY 576 -AGCAGCTGGATCTTCTCCAGCAGAGATAAACAAGAGAGTTTCCCTGACAGAAAGAT 634
Db 497 ySerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerSerThrSerSer-- 516
QY 635 GAAATATCTGCAGAGAGACATTGGGAGAGCTGGTGAACCACTTGCTGATTTCAACAGCA 694
Db 517 -----GluSerLeuAsnSerSerAlaProArgGlyGluAr 528
QY 695 ATCTACATTTCAAGACTTGGGAGTATTCTTAAGAAAAACCTTCAGAGCAAAATTCAG 754
Db 528 gSerThrLeuGlu-Asp-----SerArgGluProGlyGluAlaLeuGlyAspS 544
QY 755 ATGTTGCCAAGTC-----TCACCTGGGCGCTTTCCCTCTCA 790
Db 544 erSerAlaAsnAlaGluAspArgThrSerGlyValProSerLeuGlyThrHisThrLeuA 564
QY 791 CTCCT-----GATGACACCTCTCTCAATCAATTTCTCGAATAATACACTCAACG 838
Db 564 laThrValThrGlyAsnGlyGluArgThrLeuArgSerVal-----ThrLeuThrA 581

[illegible]

| | | | | |
|----|------|---|-----------|------|
| 27 | 1708 | CGAGACATCTGGCCCATCATGAGTACAGACAGCTCAGCGAA | -----TATG | 1810 |
| | | | | |
| Db | 879 | ysSerGlnSerThrProHisGlnGluIuIyavallieThrGluSerLySerProSerLeuV | 899 | |
| | | | | |
| Qy | 1811 | TTTTGTGTCGAGATCATTTCTTGGAGGATACCATCTGTCGTCACGCTTTACGATATCA | 1870 | |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 4, 2004, 18:45:37 ; Search time 112 Seconds
(without alignments)
12556.078 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taacaagaagggttatcct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

| |
|---------------------------|
| Xgapop 10.0 , Xgapext 0.5 |
| Ygapop 10.0 , Ygapext 0.5 |
| Fgapop 6.0 , Fgapext 7.0 |
| Delop 6.0 , Delext 7.0 |

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 1619484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_spool_p/US10007270/runat_04032004_160718_5125/app_query.fasta_1.3527
-DB=Published Applications AA -QFMT=faстан -SUFFIX=xapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR_MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10007270_@CGN 1 1 22 @runat_04032004_160718_5125
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

| | |
|-----|---|
| 1: | /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.* |
| 2: | /cgn2_6/prodata/2/pubpaa/US06_PCT_NEW_PUB.pep.* |
| 3: | /cgn2_6/prodata/2/pubpaa/US06_PUB.pep.* |
| 4: | /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.* |
| 5: | /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.* |
| 6: | /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.* |
| 7: | /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.* |
| 8: | /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.* |
| 9: | /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.* |
| 10: | /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.* |
| 11: | /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.* |
| 12: | /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep.* |
| 13: | /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.* |
| 14: | /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.* |
| 15: | /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.* |
| 16: | /cgn2_6/prodata/2/pubpaa/US10D_PUBCOMB.pep.* |
| 17: | /cgn2_6/prodata/2/pubpaa/US10E_PUBCOMB.pep.* |
| 18: | /cgn2_6/prodata/2/pubpaa/US10F_PUBCOMB.pep.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| ----- | | | | | |

| | | | | | | |
|----|--------|------|------|----|---------------------|--------------------|
| 1 | 4073 | 69.3 | 797 | 13 | US-10-007-270-2 | Sequence 2, Appli |
| 2 | 3862 | 65.7 | 771 | 13 | US-10-007-270-28 | Sequence 28, Appli |
| 3 | 3611 | 61.5 | 719 | 13 | US-10-007-270-4 | Sequence 4, Appli |
| 4 | 2504.5 | 42.6 | 798 | 13 | US-10-007-270-9 | Sequence 9, Appli |
| 5 | 1298.5 | 22.1 | 466 | 13 | US-10-007-270-11 | Sequence 11, Appli |
| 6 | 1038 | 17.7 | 198 | 13 | US-10-007-270-6 | Sequence 6, Appli |
| 7 | 917 | 15.6 | 1241 | 13 | US-10-007-270-17 | Sequence 17, Appli |
| 8 | 841 | 14.3 | 185 | 13 | US-10-007-270-15 | Sequence 15, Appli |
| 9 | 820.5 | 14.0 | 1069 | 13 | US-10-007-270-24 | Sequence 24, Appli |
| 10 | 420.5 | 7.2 | 432 | 13 | US-10-007-270-19 | Sequence 19, Appli |
| 11 | 189.5 | 3.2 | 1255 | 10 | US-09-996-069-10 | Sequence 10, Appli |
| 12 | 189.5 | 3.2 | 1255 | 14 | US-10-171-311-158 | Sequence 158, App |
| 13 | 189.5 | 3.2 | 1255 | 14 | US-10-177-293-311 | Sequence 311, App |
| 14 | 188.5 | 3.2 | 515 | 14 | US-10-097-340-212 | Sequence 212, App |
| 15 | 188.5 | 3.2 | 515 | 14 | US-10-171-311-156 | Sequence 156, App |
| 16 | 185.5 | 3.2 | 508 | 14 | US-10-057-136-20 | Sequence 20, Appli |
| 17 | 185.5 | 3.1 | 475 | 15 | US-10-417-312-1 | Sequence 1, Appli |
| 18 | 169 | 2.9 | 67 | 13 | US-10-007-270-13 | Sequence 13, Appli |
| 19 | 161.5 | 2.7 | 1905 | 15 | US-10-259-194A-86 | Sequence 86, Appli |
| 20 | 160.5 | 2.7 | 3507 | 15 | US-10-369-493-5784 | Sequence 5784, Ap |
| 21 | 151.5 | 2.6 | 1161 | 14 | US-10-017-161-2398 | Sequence 2398, Ap |
| 22 | 151.5 | 2.6 | 1161 | 15 | US-10-292-798-2040 | Sequence 2040, Ap |
| 23 | 150.5 | 2.6 | 1066 | 14 | US-09-881-752A-166 | Sequence 166, App |
| 24 | 150 | 2.6 | 1163 | 9 | US-10-029-386-32403 | Sequence 32403, A |
| 25 | 148.5 | 2.5 | 947 | 14 | US-10-293-822-1 | Sequence 1, Appli |
| 26 | 148.5 | 2.5 | 947 | 14 | US-10-177-293-86 | Sequence 86, Appli |
| 27 | 148 | 2.5 | 1490 | 15 | US-10-435-696-31 | Sequence 31, Appli |
| 28 | 148 | 2.5 | 1367 | 9 | US-09-801-368-108 | Sequence 108, App |
| 29 | 147 | 2.5 | 2409 | 14 | US-10-177-293-90 | Sequence 90, Appli |
| 30 | 146.5 | 2.5 | 630 | 14 | US-10-207-655-69 | Sequence 69, Appli |
| 31 | 145 | 2.5 | 1957 | 15 | US-10-369-493-2070 | Sequence 2070, Ap |
| 32 | 145 | 2.5 | 1431 | 14 | US-10-102-555-2 | Sequence 2, Appli |
| 33 | 144.5 | 2.5 | 1881 | 14 | US-10-032-585-7646 | Sequence 7646, Ap |
| 34 | 144 | 2.5 | 321 | 9 | US-09-925-301-861 | Sequence 861, App |
| 35 | 143 | 2.4 | 2344 | 9 | US-09-815-242-12713 | Sequence 12713, A |
| 36 | 141.5 | 2.4 | 3562 | 15 | US-10-341-434-109 | Sequence 109, App |
| 37 | 141 | 2.4 | 529 | 14 | US-10-046-420-2 | Sequence 2, Appli |
| 38 | 140.5 | 2.4 | 783 | 14 | US-10-100-957A-176 | Sequence 176, App |
| 39 | 140.5 | 2.4 | 1259 | 14 | US-10-032-585-7120 | Sequence 7120, Ap |
| 40 | 140.5 | 2.4 | 726 | 15 | US-10-012-697-1480 | Sequence 1480, Ap |
| 41 | 139.5 | 2.4 | 1928 | 15 | US-10-369-493-22025 | Sequence 22025, A |
| 42 | 139.5 | 2.4 | 1052 | 15 | US-10-094-749-2050 | Sequence 2050, Ap |
| 43 | 139 | 2.4 | 1481 | 9 | US-09-371-900-40 | Sequence 40, Appli |
| 44 | 139 | 2.4 | 1481 | 9 | US-09-924-417-60 | Sequence 60, Appli |
| 45 | 139 | 2.4 | 1481 | 9 | | |

ALIGNMENTS

RESULT 1
US-10-007-270-2
Sequence 2, Application US/10007270
Publication No. US20020160954A1
GENERAL INFORMATION:
APPLICANT: Hageman, Gregory S.
APPLICANT: Kuehn, Markus H.
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
FILE REFERENCE: 020618-000120US
CURRENT APPLICATION NUMBER: US/10/007,270
PRIOR FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 09/430,195
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: US 09/183,972
PRIOR FILING DATE: 1998-10-29
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 797
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Human IPM 150 amino acid sequence, isoform A

US-10-007-270-2

Alignment Scores:

Pred. No.: 0
 Score: 4073.00
 Percent Similarity: 97.31%
 Best Local Similarity: 97.19%
 Query Match: 69.33%
 DB: 13
 Length: 797
 Matches: 796
 Conservative: 1
 Mismatches: 0
 Indels: 22
 Gaps: 1

US-10-007-270-1 (1-3330) x US-10-007-270-2 (1-797)

| | | | |
|----|------|--|------|
| QY | 131 | ATGATTTTGGAACTAGAGAGCTATTTTCTTTTGGATTTTCTCCAGTTCAGGA | 190 |
| DB | 1 | MetTyrLeuGluThrArgAlaValPheThrPheLeuGlnValGlnGly | 20 |
| QY | 191 | ACCAAGATATCTCCATTACATATACCATCTGAACTAAAGACATAGACATCCCCCA | 250 |
| DB | 21 | ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAspAsnProPro | 40 |
| QY | 251 | AGAAATGAAACCACTGAAAGTACTGAAATAATCTCAAAATGTCAACTATGAGAGGAATA | 310 |
| DB | 41 | ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgile | 60 |
| QY | 311 | TTGATTTTGGCAAGCATCGAACAAGATCGCATTTTCCCAAGCGGGTAAAGTC | 370 |
| DB | 61 | PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal | 80 |
| QY | 371 | TGTCACAGGANTCCATGAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA | 430 |
| DB | 81 | CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrArgLeuArg | 100 |
| QY | 431 | GTGTGTCAGAACGACATGGAAGCATATCGGATCTTTCTGAGATCGATCCCTGACACA | 490 |
| DB | 101 | ValCysGlnGlnAlaValTrpGluAlaTyrArgIlePheLeuAspArgileProAspThr | 120 |
| QY | 491 | GGGGAATATCAGACTGGGTGAGCATCTGCCAGAGAGACCTTCTGCTCTCTTGGCATTT | 550 |
| DB | 121 | GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle | 140 |
| QY | 551 | GGAAAACTTCAGCAATTCAGGAGCACTGGATCTTCTCCAGCAGAGATAAAACAG | 610 |
| DB | 141 | GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln | 160 |
| QY | 611 | AGAAATTTCCCTGACAGAAAAGATGAATATCTGCAGAGAGACATTTGGAGAGCTGGT | 670 |
| DB | 161 | ArgSerPheProAspArgLysAspGluIleSerAlaGluLysThrLeuGluProGly | 180 |
| QY | 671 | GAACCAATGTCTATTCACAGCAATCTACATTTCAAAGACTTGGGCAGTATTCTAAGAA | 730 |
| DB | 181 | GluThrIleValIleSer--Thr----- | 187 |
| QY | 731 | AACCTTCAGAGAGCAAAATTCAGAGATGTGGACGCTCTCACTGGGCTTTCCTCTCA | 790 |
| DB | 188 | -----AspValAlaAsnValSerLeuGlyProPheProLeu | 200 |
| QY | 791 | CTCTGATGACACCTCTCTCAATGAATTTCTCGATATATACACTCAACACACACAGATGC | 850 |
| DB | 200 | hrProAspAspThrLeuLeuAsnGluIleLeuAspAsnThrLeuAsnAspThrLysMetP | 220 |
| QY | 851 | CTACACAGAAAGAGAAACGAATTCGTGTGTGGAGGAGCAGAGGTGGAGCTCAGCG | 910 |
| DB | 220 | roThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV | 240 |
| QY | 911 | TCTCTCTGTAACCCAGAGTTTCAAGGTCAGAGCTCGCTGACTCCAGTCCCAATATTACC | 970 |
| DB | 240 | alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG | 260 |
| QY | 971 | AGAGCTAGCAGGAAGTCCCACTTCAGATGCAAAAGATATTAAAGAACTTCCAGAT | 1030 |
| DB | 260 | InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP | 280 |
| QY | 1031 | TCAAAAAATCCATGTGTAGGATTTAGACAAAGAGAAAAAGATGGCTCAAGCTCCA | 1090 |

| | | | |
|----|------|--|------|
| DB | 280 | helyLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert | 300 |
| QY | 1091 | CAGAGATGCAACTTACGCCCATCTTTAAGACACACAGTGCAGAGCAAAAGCCCTGCAA | 1150 |
| DB | 300 | hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlas | 320 |
| QY | 1151 | GTGACCTCTCTCTTTTGTATCCAAACAAATTTGAAGTGAAGGAGTATCATCGAACCA | 1210 |
| DB | 320 | erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM | 340 |
| QY | 1211 | TGGAGGACAGACACCAACCAATATCTATCTCAGCTACAGACCTCAAAAGGCTGATCA | 1270 |
| DB | 340 | etGluGlnAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIles | 360 |
| QY | 1271 | GCAAGCACTAGAGAGAGAACAACTCTTTGGATGTGGGACAACTCAGTTCAGTCAATGA | 1330 |
| DB | 360 | erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI | 380 |
| QY | 1331 | TTGCTGATCATCTCCAGCTTTGTCTCTGACACCAATCAGAGCTGCCACATCTTTTG | 1390 |
| DB | 380 | IeAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA | 400 |
| QY | 1391 | CTGTATTAACAGAGATGTCTTCTTGTAGTCCAGAACTTCTCTCTGTGGAAACCCAGCTG | 1450 |
| DB | 400 | IaValIleThrGluAspAlaThrLeuSerProGluLeuProProValGluProGlnLeuG | 420 |
| QY | 1451 | AGAGCTGGACGAGCAGAGCATGTCTTCTACGACACTTCTTGTCTCCACCTGTATGG | 1510 |
| DB | 420 | IuThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProProAlaMetA | 440 |
| QY | 1511 | CTCTACCTCTCTCTCAGAGCTCCACTTCTTTTATGGCATCAAGCATCTTCTCTCTGA | 1570 |
| DB | 440 | IaserThrSerLeuSerGluAlaProPhePheMetAlaSerSerIlePheSerLeuT | 460 |
| QY | 1571 | CTGATCAAGCAGCAGACATCAATGCGCACTGACCACTGGCTCTGGGAAATTCACATCCAC | 1630 |
| DB | 460 | hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT | 480 |
| QY | 1631 | CAATCCCTCAGCATGATTTATCTCCATCAGCCAACTGGCTCTGGGAAATTCACATCCAC | 1690 |
| DB | 480 | hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp | 500 |
| QY | 1691 | CTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCGAAGATATGTCAGACACCTAG | 1750 |
| DB | 500 | roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAspMetValArgHisLeuA | 520 |
| QY | 1751 | ATGAATGGATCTGTCTGACACTCTCTGCCCATCTGAGGTACAGAGCTCAGAGCAATATG | 1810 |
| DB | 520 | epGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV | 540 |
| QY | 1811 | TTTCTGTCCAGATCATTTCTGGAGATPACCATCTCTGTCTCAGCTTTACAGTATATCA | 1870 |
| DB | 540 | alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT | 560 |
| QY | 1871 | CCACTAGTTCTATACCATTTGCCCCAGGCGGAGAGCTGTGTAGTGTCTTCTGCTCTGC | 1930 |
| DB | 560 | hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA | 580 |
| QY | 1931 | GTGTCTTCAATGCGCTTCTCCACGACCTGTTCACCAAGAGCTCTCTGAGTACCGAG | 1990 |
| DB | 580 | rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA | 600 |
| QY | 1991 | CTCTGGAGCACAATTCACAGCTGTGTGTTCATATCTACGATCCAATCTTACAGGAT | 2050 |
| DB | 600 | IaLeuGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP | 620 |
| QY | 2051 | TTAAGCAACTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 2110 |
| DB | 620 | helyGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL | 640 |
| QY | 2111 | AGTTTGTAGTCTGTGGCGTATTAACCTCAAGAGCTGTGCAGCGGTCTTGGAGGAT | 2170 |

Db 640 ysPheAlaLysSerValProTyAsnLeuThrLysAlaValHisGlyValLeuGluAsp 660
 QY 2171 TTGCTTCTGTCGACCCCAACAACTCCATCTGGAATAGACAGCTACTCTCAACATTG 2230
 Db 660 heArgSerAlaAlaGlnGlnLeuHisLeuGluLeuLeuLeuLeuLeuLeuLeuLeu 680
 QY 2231 AACGAGCTGATCAAGCAGATCCCTGCAAGTTCCTGCGCTCGCGCAATTTCCCAATGTG 2290
 Db 680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
 QY 2291 TAAAGAACGAGCTGAGAAAGCGAGTTCCTGCAACCAAGCAGATATCAGACGAG 2350
 Db 700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyAspSerGlnG 720
 QY 2351 GGAGCTGACAGCTGGAACCGAGCTCTGTGCGCTCGCACAAAGGAATGGCAGGTCC 2410
 Db 720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 740
 QY 2411 TCCAGGGAAGCGAGCTCCATGCGAGTTCGAGATCACTCTGAAATCAGCATCAAAA 2470
 Db 740 eugInGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyLys 760
 QY 2471 CTAGTGTAAAGTTCCAAATCAACAAATAACAGGTAAATCAAGTAAAGAAATTCG 2530
 Db 760 hrSerValLysLysPheGlnAsnGlnGlnAsnLysValLysSerLysArgAsnSerG 780
 QY 2531 AATTACTACCTAGATATGAGAAATTTAACATCAAGATTTGGAAGGAAAT 2583
 Db 780 luLeuLeuThrValGluTyGluGluPheAsnHisGlnAspTrpGluGlyAsn 797

RESULT 2

US-10-007-270-28
 ; Sequence 28, Application US/10007270
 ; Publication No. US20020160954A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hageman, Gregory S.
 ; APPLICANT: Kuehn, Markus H.
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
 ; FILE REFERENCE: 020618-000120US
 ; CURRENT APPLICATION NUMBER: US/10/007,270
 ; CURRENT FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: US 09/430,195
 ; PRIOR FILING DATE: 1999-10-29
 ; PRIOR APPLICATION NUMBER: US 09/183,972
 ; PRIOR FILING DATE: 1998-10-29
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 771
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; OTHER INFORMATION: Human IPM 150 isoform A variant cDNA sequence
 US-10-007-270-28

Alignment Scores:
 Pred. No.: 3,496-312 Length: 771
 Score: 3862.00 Matches: 769
 Percent Similarity: 96.74% Conservative: 2
 Best Local Similarity: 96.49% Mismatches: 0
 Query Match: 65.74% Indels: 26
 DB: 13 Gaps: 1

US-10-007-270-1 (1-3330) x US-10-007-270-28 (1-771)

QY 131 ATGATTGGAACTAGAGAGCTATTTTGTGTTTGTGTTTCTCCAGTTCAAGGA 190
 Db 1 MetTyLeuGluThrArgArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCCATTACATATACCTTTCTGAACTAAAGACATAGACAATCCCCA 250
 Db 21 ThrLysAspIleSerIleAsnIleTyHisSerGluThrLysAspIleAsnProPro 40

QY 251 AGAATGAACAACTAGAGTACTCAAAATATGTACAAATGTCACTATGAGACGAATA 310
 Db 41 ArgAsnGluThrThrGluSerThrGluLysMetTyLysMetSerThrMetArgArgile 60
 QY 311 TTGATTTGGCAAGCATCGAACCAAAAGATCCGATTTTCCCAACGGGGGTAAAGTC 370
 Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
 QY 371 TGTCCACAGGAATCCATGAACAGATTTAGACAGCTCTTCAAGCTTATTATAGATTGAGA 430
 Db 81 CysProGlnGlnGlnMetLysGlnIleLeuAspSerLeuGlnAlaTyTyArgLeuArg 100
 QY 431 GTGTGTCCAGGACAGTATGGAGAGCATATCCGATTTCTGATCGCATCCCTGACACA 490
 Db 101 ValCysGlnGlnAlaValTrpGluAlaTyArgIlePheLeuAspArgileProAspThr 120
 QY 491 GGGGAATATCAGGACTGGGTCCAGCATCTCCAGCAGGAGACCTTCTGCTCTTTGACATT 550
 Db 121 GlyGluTyGlnAspTrpValSerIleCysGlnGlnGlnThrPheCysLeuPheAspIle 140
 QY 551 GGAATAAATCTCAGCAATTCGAGGAGCAGCTGGATCTTCCAGCAGAGATATAAAGCAG 610
 Db 141 GlyLysAsnPheSerAsnSerGlnGlnHisLeuAspLeuLeuGlnGlnArgIleLysGln 160
 QY 611 AGAAGTTTCCCTGACAGAAAGATGAATATCTCAGCAGAGAGACATTTGGGAGCCTGGT 670
 Db 161 ArgSerPheProAspArgLysAspGluLysSerAlaGluLysThrLeuGlyGluProGly 180
 QY 671 GAAACCATTTGTCATTCAACAGCAATCTACATTTTCAAAGACTTTGGCGCATTTCTAAGAA 730
 Db 181 GluThrIleValIleSer--Thr----- 187
 QY 731 AACCTCAGAGAGCAATTCAGATTTGTCACAGCTCTCACTTGGGCTTTCCCTCTCA 790
 Db 188 -----AspValAlaAsnValSerLeuGlyProPheProLeu 200
 QY 791 CTCTCTGACACACCTCTCTCAATGAATTCGATATATCACTCAACGACACCAAGTCG 850
 Db 200 hrProAspAspThrLeuLeuAsnGluIleLeuAspIleThrLeuAsnAspThrLysMetP 220
 QY 851 CTACACAGAAAGAGAAACAGAAATTCGTGTTGGAGGAGCAGAGGTGGAGCTCAGCG 910
 Db 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 240
 QY 911 TCTCTCTGTAAACCAAGTTCAAGCAGAGCTCCCTGACTCCCGAGTCCCGCATATTACC 970
 Db 240 aSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyTyG 260
 QY 971 AGGAGCTAGCAGGAAAGTCCCAACTTCAGATGCAAAAGATATTTAAGAAACTTCCAGGAT 1030
 Db 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyP 280
 QY 1031 TCAAAAATCCATGTGTTAGATTAGCAAAAGAAAGAAAGATGGCTCAAGCTCCA 1090
 Db 280 heLysLysIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 300
 QY 1091 CAGAGATCAACTTACGCGCATCTTAAAGAGACACAGCTGCAAGCAAAAGCCCTGCAA 1150
 Db 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAla 320
 QY 1151 GTGACCTCTCTGTTTGTGTTTCAACAAATTTGAAGTGAGAGATCTATCATGGACCA 1210
 Db 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyHisGlyThrM 340
 QY 1211 TGGAGGAGGACAGCAACCAAGATCTATCTCAGCTACAGCTCAAGCTCAAAAGCTGATCA 1270
 Db 340 etGluGluAspLysGlnProGluIleTyLeuThrAlaThrAspLeuLysArgLeuLys 360
 QY 1271 GCAAAGCACTAGAGGAGAAACAATTTTGGATGTGGGACAAATTCATGTTCACTGATGAAA 1330
 Db 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380
 QY 1331 TTGCTGGATCATCTGCCAGCCTTTTGTCTCTGACACCAATCAGAGCTGCCCATCTTTTG 1390

Db 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400
QY 1391 CTGTTATTAACAGAGGAGTCTACTTTGAGTCCAGAACTTCTCTCTGTTGTAACCCAGCTTG 1450
Db 400 laValleThrGluAaspAlaThrLeuSerProGluLeuProProValGluProGlnLeuG 420
QY 1451 AGACAGTGGAGGAGCAGAGCATGCTTACTACGACACTTCTGGTCTCCACCTGATGG 1510
Db 420 luThrValAaspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 440
QY 1511 CCTTACCTCCCTGTCTCAGAGCTCCACCTTCTTTATGSCATCAACACTCTCTCTGA 1570
Db 440 laSerThrSerLeuSerGluAlaProProPhePheAlaSerSerIlePheSerLeuT 460
QY 1571 CTGATCAAGCACCACAGATCAATGGCCACTGACCAGACAATGCTAGTACCAGGGCTCA 1630
Db 460 hrAaspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480
QY 1631 CCATCCCCACCATGATTCTTCCATTCAGCACTGCTGGTCTGGGAATTCACATCCAC 1690
Db 480 hrIleProThrSerAspTrpSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500
QY 1691 CTGCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCGAAGATATGCTCAGACACCTAG 1750
Db 500 roAlaSerSerAspAspSerArgSerSerAlaGlyGlyGluAaspMetValArgHisLeuA 520
QY 1751 ATGAATGAGATCTGTCTGACACTCTGCCCCATCTGAGGTACCAGAGCTCAGCAATATG 1810
Db 520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrv 540
QY 1811 TTTCTGTCCACATCATTTCTTGAGGATACACTCTCTCTCTCAGCTTTACAGTATATCA 1870
Db 540 alSerValProAspHisPheLeuGluAaspThrThrProValSerAlaLeuGlnTyrlleT 560
QY 1871 CCACTAGTTCTATGACCATGCCCCCAAGGGCCGAGAGCTGGTAGTGTCTCAGCTCTC 1930
Db 560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580
QY 1931 GTGTCTGTAACATGGCTTCTCCACGACCTGTGTCACAGAGCTCTCTGGAGTACCAG 1990
Db 580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrvArgA 600
QY 1991 CTCTGGAGCAACAATTCAACAGCTGCTGTTCCATATCTACGATCCATCTTACAGGAT 2050
Db 600 laLeuGluGlnPheThrGlnLeuLeuValProTyrvLeuArgSerAsnLeuThrGlyp 620
QY 2051 TTAAGCACTTGAATACTTAACTTCAGAAACGGGAGTGTGATTGTGAATAGCAAAATGA 2110
Db 620 heLysGlnLeuGluLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
QY 2111 AGTTTGCTAAGTCTGCGTATACCTTACCAAGCTGTGCAAGGGTCTTGGAGGATT 2170
Db 640 ysPheAlaLysSerValProTyrvAsnLeuThrLysAlaValHisGlyValLeuGluAasp 660
QY 2171 TTGTTTCTGTCGAGCCCAACAATCCATCTGGAATAGACAGCTCTCTCTCAACATTG 2230
Db 660 heArgSerAlaAlaAlaGlnLeuHisLeuGluIleAspSerTyrvSerLeuAnIleG 680
QY 2231 AACAGCTGATCAAGCAGATCCCTGCAAGTCTCTGCGCTGGCGGAATTTGCCCAATGTG 2290
Db 680 luProAlaAspGlnAlaAaspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysv 700
QY 2291 TAAAGCAACAGGACTGAGGAGCGGAGTGTGCTGCAACAGGATATGACGCCAGG 2350
Db 700 alLysAsnGluuThrGluGluAlaGluCysArgCysLysProGlyTyrvAspSerGlnG 720
QY 2351 GGAGCTGGAGCTGTGGAAACAGGCTCTGTGCGCTGGCACAAGGATCGAGTCC 2410
Db 720 lYserLeuAspGlyLeuGluProGlyLeuCysGlyLeuAlaGlnArgAsnAlaArgSer 739
QY 2411 TCCAGGGAAGGAGTCCATGCGAGGTGCGAGATCACTCTGAAAATCAAGCATCAAAA 2470
||| :|||

Db 740 SerArgGluArgGluLeuHisAla--ValProAspHisSerGluAsnGlnAlaTyrlYst 759
QY 2471 CTAGTGTAAAAAGTTCCAAAAATCAACAAATAACCAAGG 2509
Db 759 hrSerVal-LysSerSerLysIleAsnLysIleThrArg 771
RESULT 3
US-10-007-270-4
; Sequence 4, Application US/10007270
; Publication No. US20020160954A1
; GENERAL INFORMATION:
; APPLICANT: Hageman, Gregory S.
; APPLICANT: Kuehn, Markus H.
; TITLE OF INVENTION: University of Iowa Research Foundation
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
; FILE REFERENCE: 020618-000120US
; CURRENT APPLICATION NUMBER: US/10/007,270
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 09/430,195
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 09/183,972
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Human IPM 150 amino acid sequence, isoform B
US-10-007-270-4
Alignment Scores:
Pred. No.: 2,6e-291 Length: 719
Score: 3611.00 Matches: 717
Percent Similarity: 87.67% Conservative: 1
Best Local Similarity: 87.55% Mismatches: 1
Query Match: 61.46% Indels: 100
DB: 13 Gaps: 2
US-10-007-270-1 (1-3330) x US-10-007-270-4 (1-719)
QY 131 ATGTATTGGAACTAGAAAGCTATTGTTTGGATTTTCTCCAGTTCAGGA 190
|||
Db 1 MetTyvLeuGluThrArgAgaIallePheValPheIlePheLeuGlnValGlnGly 20
QY 191 ACCAAGATATCTCCATTAACATATACCATCTCGAACTAAAGACATAGACAATCCCCCA 250
|||
Db 21 ThrLys----- 22
QY 251 AGAAATGAAACAACATGAAAGTACTGAAAAAATGTACAAAATGTCAACTATGAGACGAATA 310
|||
Db 22 ----- 22
QY 311 TTCGATTGGCAAGCATCGAACAAGATCCGCATTTTCCCAACGGGGGTTAAAGTC 370
|||
Db 22 ----- 22
QY 371 TGTCACAGGAATCCATGAAACAGATTTTAGACAGCTTTCAAGCTTATTATAGATTGAGA 430
|||
Db 22 ----- 22
QY 431 GTGTGTGAGGAAGCAGTATGGAAGCATATCGGATCTTCTGGATCCATCCCTGACACA 490
|||
Db 23 ValCysGlnGluAlaValIlePheGluAlaTyvArgIlePheLeuAspArgIleProAspThr 42
QY 491 GGGAATATCAGGATCGGTGAGCATCTGCGACGAGAGACTCTCTGCTCTTTCGATTT 550
|||
Db 43 GlyTyvGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 62
QY 551 GGAAAAATCTCAGCAATTTCCACAGAGCACTCGATCTTCTCCAGCAGAGAAATAAACA 610
|||
Db 63 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnArgIleLysGln 82

QY 611 AGAAGTTTCCTGACAGAGAAAGATGAATATCTGCAGAGAGAGCAATTTGGAGAGCCCTGGT 670
 Db 83 ArgSerPheProAspArgLysAspGluLeuSerAlaGluLeuThrLeuGluProGly 102
 QY 671 GAAACCATTTGCAATTCACAGCAATCTACATTTCAAAGACTTGGGCGAGTATCTAAGAA 730
 Db 103 GluThrIleValIleSer--Thr----- 109
 QY 731 AACCTTCAGAGAGCAAAATTCAGAGATGTGCAACAGCTCACTTGGGCGCTTCCCTCTCA 790
 Db 110 -----AspValAlaAsnValSerLeuGlyProPheProLeu 122
 QY 791 CTCCTGATGACACCCCTCCTCAATGAATTCGATTAATACACTCAAGCAGCACCAAGATGC 850
 Db 122 hrProAspThrLeuLeuAsnGluLeuAsnGluLeuAsnThrLeuAsnAspThrLysMetp 142
 QY 851 CTACACAGAAAGAGAAACAGAATTCGCTGTGTGGAGGAGCAGAGGTCAGAGCTCAGCG 910
 Db 142 roThrGluArgGluThrGluPheAlaValLeuGluGluGluArgValGluLeuSerV 162
 QY 911 TCTCTCTGTAACCAAGAGTTCAAGGACAGAGCTCGCTGACTCCAGTCCCATATTACC 970
 Db 162 alSerLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG 182
 QY 971 AGGAGTAGCAGGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAACTTCAGAGAT 1030
 Db 182 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLysLeuProGlyp 202
 QY 1031 TCMAAAAANTCCATGTGTAGGATTTAGACCAAGAAAGAAAGATGCTCAAGCTCCA 1090
 Db 202 helyslysiIleHisValLeuGlyPheArgProLysLysGluLysAspGlySerSerSert 222
 QY 1091 CAGAGATGCAACTTACGCCATCTTTAAGAGACACAGTGCAGAGCAAAAGCCCTGCAA 1150
 Db 222 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAlas 242
 QY 1151 GTGACTCTGCTCTTTGATTCACAAATTAAGAGTGAAGAGTCTATCATGGAACCA 1210
 Db 242 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 262
 QY 1211 TGAGGAGCAGCAGCAACCAATCTATCTCAGCTACAGCTCAAAAGGCTGATCA 1270
 Db 262 etGluLysAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuIleS 282
 QY 1271 GCAAGCCTAGCAGAGAAACAATCTTGGATGTGGGCAATTCAGTCACTGATGAAA 1330
 Db 282 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 302
 QY 1331 TTCTGGATCACTGCCAGCCTTTGCTCTGACACCCCAATCAGAGCTGCCACATCTTTTG 1390
 Db 302 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 322
 QY 1391 CTCTTATAACAGAGATGCTACTTTGAGTCCAGAACTTCTCTGTTGAACCCAGCTTG 1450
 Db 322 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 342
 QY 1451 AGACAGTGGACGAGAGACATGCTCTACCTGACACTCTCTGCTGCTCCAGCTGTATGG 1510
 Db 342 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProAlaMetA 362
 QY 1511 CTTCTACCTCTCTGTGAGAGCTCCACTTTCTTTATGTCATCAAGCATCTCTCTGGA 1570
 Db 362 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerIlePheSerLeu 382
 QY 1571 CTGATCAAGGCACACAGATACATGCGCACTGACAGCAATGCTAGTACCAGGGCTCA 1630
 Db 382 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeu 402
 QY 1631 CCATCCCCACAGTATTATCTGCAATCAGCACTGGCTCTGGGAATTCACATCCAC 1690
 Db 402 hrIleProThrSerAspTyrSerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 422

QY 1691 CTGCATCTTTCAGATCAGACGCCGATCAAGTSCAGGTGGGAGATATGTCAGACACTAG 1750
 Db 422 roAlaSerSerAspAspSerArgSerAlaGlyGlyGluAspMetValArgHisLeuA 442
 QY 1751 ATGAAATGGATCTGTCTCAGACTCCCTGCCCCCATCTGAGGTACCGAGCTCAGCGAATATG 1810
 Db 442 epGluMetAspLeuSerAspThrProAlaProSerGluValProGlyLeuSerGluTyrV 462
 QY 1811 TTTCTGTCACGATCATTTCTTGGAGGATACCACCTCTCTGCTCAGCTTACAGTATATCA 1870
 Db 462 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIle 482
 QY 1871 CCCTAGTTCATGACCATTTGCCCCCAAGGCCGAGAGCTGTGTAGTGTCTTTCAGTCTG 1930
 Db 482 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 502
 QY 1931 GTCTTCTTAAACATGCGCTTCTCCAGACCTGTTCAACAGAGCTCTCTGAGTACCGAG 1990
 Db 502 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 522
 QY 1991 CTCTGAGCAACAATTCACACAGCTGCTGGTTCATATCTAGATCCCAATCTTACAGGAT 2050
 Db 522 laLeuGluGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyp 542
 QY 2051 TTAAGCAACTTGAATTAATTAACCTTACAGAAACGGAGTGTGATTTGAATAGCAAAATCA 2110
 Db 542 helysleuGluGluLeuLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 562
 QY 2111 AGTTTGTCTAGTCTGTGCGGTAACTTCACTCACCAGGCTGTCCACGGGCTTGGAGGAT 2170
 Db 562 yspPheAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAsp 582
 QY 2171 TTCTGCTGTCTGAGGCCCAACAATCCATCTGGAATAGACAGCTACTCTCTCAACATG 2230
 Db 582 heArgSerAlaAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 602
 QY 2231 AACAGCTGATCAAGCAGATCCCTGCAAGTTCCTGCGCTCGCGGCAATTTGCCAATGTG 2290
 Db 602 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 622
 QY 2291 TAAAGAACAGACGACTGAGAAAGCGAGTGTGCTGCAACACAGGATATGACCCAGG 2350
 Db 622 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGlnG 642
 QY 2351 GAGCCTTGACCGCTCTGAGACAGGCTCTGTGCGCTCTGCGCAACAGGAATGCCAGTCC 2410
 Db 642 lysSerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 662
 QY 2411 TCAGAGGAAAGCGAGCTCCATGCAAGTTCGCGATCACTCTGAAATCAAGCATACAAA 2470
 Db 662 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrLysP 682
 QY 2471 CTAGTCTTAAAGTTCAAAATCAACAAAATAACAGGTAAATCAAGTAAGAAATCTG 2530
 Db 682 hrSerValLysLysPheGlnAsnGlnGlnAsnLysValIleSerLysArgAsnSerG 702
 QY 2531 AATTACTGACCTGATATATGAAATTAACCATCAAGATTCGGAAGGAAAT 2583
 Db 702 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGluYan 719

RESULT 4

US-10-007-270-9
 ; Sequence 9, Application US/10007270
 ; Publication No. US20020160954A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hageman, Gregory S.
 ; APPLICANT: Kuehn, Markus H.
 ; APPLICANT: University of Iowa Research Foundation
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
 ; FILE REFERENCE: 020618-00012005
 ; CURRENT APPLICATION NUMBER: US/10/007,270
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: US 09/430,195

; PRIOR FILING DATE: 1999-10-29
 ; PRIOR APPLICATION NUMBER: US 09/183,972
 ; NUMBER OF SEQ. IDS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ. ID NO. 9
 ; LENGTH: 798
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 ; FEATURE:
 ; OTHER INFORMATION: Mouse IPM 150 amino acid sequence, isoform A
 US-10-007-270-9

Alignment Scores:

Pred. No.: 3.17e-199 Length: 798
 Score: 2504.50 Matches: 527
 Percent Similarity: 73.13% Conservative: 80
 Best Local Similarity: 63.49% Mismatches: 180
 Query Match: 42.63% Indels: 43
 DB: 13 Gaps: 10

US-10-007-270-1 (1-3330) x US-10-007-270-9 (1-798)

```

QY 131 ATGCTATTGGAACTAGAGAGCTATTTTGTGTTTTCCTCAAGTTCAGGA 190
Db      |||      |||      |||      |||      |||      |||      |||
1 MetAsnPheGlnIleLysHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20

QY 191 ACCAAAGATATCTCCATTAACATATACCATCTCTGAACTTAAGACATAGACCAATCCCCCA 250
Db      |||      |||      |||      |||      |||      |||      |||
21 IleLysAspThrSerIleLysIlePheSerSerGluIleLysAsnIleAspLysThrPro 40

QY 251 AGAATGAACAACACTGAAGTACTGAAAAATGTACAAATGTCAAACTATGAGCGAATA 310
Db      |||      |||      |||      |||      |||      |||      |||
41 ArgIleGluThrIleGluSerThrSerValHisLysValSerThrMetLysArgIle 60

QY 311 TTCCGATTGGCAAGCATCGAAACAAAGATCCGATTTTCCCAACGGGGGTTAAAGTC 370
Db      |||      |||      |||      |||      |||      |||      |||
61 PheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro---AlaAsnIle 79

QY 371 TGTCACAGGATTCATCAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTGAGA 430
Db      |||      |||      |||      |||      |||      |||      |||
80 CysProGlnGluSerLeuArgGlnIleLeuAlaSerLeuGlnGluIleLysArgLeuArg 99

QY 431 GTGTGTGACAGACAGTATGGGAAGCATATCGATCTTCTGATCGATCCCTGACACA 490
Db      |||      |||      |||      |||      |||      |||      |||
100 ValCysGlnGluValIleTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 119

QY 491 GGGGAATATCAGACTGGGTGACGATCTCCAGCAGGAGACCTTCTGCTCTTTGACATT 550
Db      |||      |||      |||      |||      |||      |||      |||
120 GluGluTyrGlnAspTrpValSerLeuCysGlnLysGluThrPheCysLeuPheAspIle 139

QY 551 GMAAAACCTTCAGCAATTCACAGGACACTGATCTTCTCCAGCAGAGATAAAGCAG 610
Db      |||      |||      |||      |||      |||      |||      |||
140 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 159

QY 611 AGAAGTTTCCCTGACAGAAAGATATATCTGACAGAGACATTTGGAGAGCCTGGT 670
Db      |||      |||      |||      |||      |||      |||      |||
160 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 179

QY 671 GAAACCATTTGTCATTTCAACAGCATCTACATTTTCAAGACTTGGGCGAGTATTTCAAGAA 730
Db      |||      |||      |||      |||      |||      |||      |||
180 GluAlaProValValPro---Thr-----186

QY 731 AACCTTCAGAGAGCAAAATTCAGATGTTGCCAGCTCTCACTTGGGCTTTCCCTCTCA 790
Db      |||      |||      |||      |||      |||      |||      |||
187 -----AspValSerArgMetSerLeuGlyProPheProLeuP 199

QY 791 CTCCTGTGACACCTCTCAATGAATTTCTCGATTAATCACTCAACACCAAGATGC 850
Db      |||      |||      |||      |||      |||      |||      |||
199 roSerAspAspThrAspLeuLysGluIleLeuSerValThrLeuLysAspIleGlnLysP 219

QY 851 CTACACCAAGAGAAACA-----GAATTGCTGTGTTGGAGGAGC 892
Db      |||      |||      |||      |||      |||      |||      |||

```

```

Db 219 roThrThrGluSerLysThrGluProIleHisValSerGluPheSer-----SerGluG 237
QY 893 AGAGGTGGAGCTCAGCGTCTCTGTGTAAACCAAGAAAGTTTCAAGCGAGAGCTCGCTGACT 952
Db      |||      |||      |||      |||      |||      |||      |||
237 IuLysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS 257
QY 953 CCCAGTCCCCTATATTACAGGAGCTAGCAGGAAAGTCCCACTTCCAGATGCAAGATAT 1012
Db      |||      |||      |||      |||      |||      |||      |||
257 erGlySerProTyrTyrGlnGluLeuValGlyGlnSerGlnLeuGlnLysIleP 277
QY 1013 TTAAGAAATCTCCAGGATTCAAAAAATCCATGTGTAGATTAGACCAAGAAAGAA 1072
Db      |||      |||      |||      |||      |||      |||      |||
277 helyLysLeuProGlyPheGlyGluIleArgValLeuGlyPheArgProLysGluG 297
QY 1073 AAGATGCTCAGCTCCACAGAGATGCAACTTACGCCCATCTTTAAGAGACACAGTGCAG 1132
Db      |||      |||      |||      |||      |||      |||      |||
297 IuAspGlySerSerThrThrGluIleGlnLeuMetAlaIlePheLysArgAspHisAlaG 317
QY 1133 AAGCAAAAAGCCCTGCAAGTACCTCCCTGCTTTTTCATTCACCAAAAATTTGAAGTGAGG 1192
Db      |||      |||      |||      |||      |||      |||      |||
317 IuAlaLysSerProAspSerHisLeuLeuSerLeuAspSerAsnLysIleGluSerGluA 337
QY 1193 AAGTCTATCTGGAACCAATGAGGAGGAGCAAGCAACCAAAATCTATCTCACAGCTACAG 1252
Db      |||      |||      |||      |||      |||      |||      |||
337 rgIleHisIleGlyValIle---GluAspLysGlnProGluThrTyrLeuThrAlaThrA 356
QY 1253 ACTCAAAAGGCTGATCACCAAGCACTAGAGGAGACAATCTTTGGATGTGGGCAAA 1312
Db      |||      |||      |||      |||      |||      |||      |||
356 spluLysLysLeuIleGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI 376
QY 1313 TTCAGTCTCACTGATGAATTTGCTGATCACTGCCAGCTTGTGGCTTCAGACCCCAATCAG 1372
Db      |||      |||      |||      |||      |||      |||      |||
376 leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA 394
QY 1373 ACTGCCCAACATCTTTTGTGTATTAACAGAGAGTCTACTTTGAGTCAGCAACTCTCTC 1432
Db      |||      |||      |||      |||      |||      |||      |||
394 spluProLysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluLeuProP 414
QY 1433 CTGTGTAAACCCAGCTTCAGACAGTGGAGGAGAGAGCATGTGCTTACCT-----1482
Db      |||      |||      |||      |||      |||      |||      |||
414 heValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS 434
QY 1483 -----GACACTTTTGGTCTCCACCTGCTATGGCTCTACTCTCTCCCTGTCAGAGCTCCAC 1537
Db      |||      |||      |||      |||      |||      |||      |||
434 erLysAspSerSerTrpSerProValSerAlaSerIleSerArgSerGluAsnLeuP 454
QY 1538 CTCTTCTTATGGATCAAGCATCTTCTCTGACTGATCAAGGACCAACAGATACAATGG 1597
Db      |||      |||      |||      |||      |||      |||      |||
454 roSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProLeuMetT 473
QY 1598 CCACTGACCAACAATGCTAGTACCAGGGCTCACCATCCCAACAGTGAATTTCTGCAA 1657
Db      |||      |||      |||      |||      |||      |||      |||
473 hrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTyrSerThrI 493
QY 1658 TCAGCCAACTGCTGCTGGGAATTTCACTCCACCTGATCTTCCAGATCAGACCGCATCA 1717
Db      |||      |||      |||      |||      |||      |||      |||
493 leArgGlnLeuProLeuGluSerSerHisTrpProLaserSerSerAspArgGluLeuI 513
QY 1718 GTCCAGGTGGCAAGATATGTCAGACACCTAGATGAATGGATCTGTCTGACATCTCTG 1777
Db      |||      |||      |||      |||      |||      |||      |||
513 leThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspThrProA 533
QY 1778 CCCACTGTGAGTACAGAGCTCAGCAATATGTTTCTCTCCAGATCATTTCTTGAGG 1837
Db      |||      |||      |||      |||      |||      |||      |||
533 laLeuSerGluIleSerGluLeuSerGlyIleAspSerAlaSerGlyLysPheLeuGluM 553
QY 1838 ATACCACTCTGCTCAGCTTTACAGTATATCACCTAGTGTCTATCACCATTGCCCCCA 1897
Db      |||      |||      |||      |||      |||      |||      |||
553 etThrThrProIleProThrValArgPheIleThrSerSerGluThrIleAlaThrL 573
QY 1898 AGGGCCGAGAGCTGTGATGTTCTTCTAGTCTGCGTGTGCTAAACATGGCTTCTTCAACG 1957
Db      |||      |||      |||      |||      |||      |||      |||
573 ysGlyGlnGluLeuValPhePheSerLeuArgValAlaAsnMetProPheSerTyrA 593

```


Db 352 ----- 352
 QY 2213 GCTACTCTCTCAACATTGAACACGAGTGTATCAAGCAGATCCCTGCAAGTCTCTGAGCTGCG 2272
 Db 353 ----- ProAlaAspGlnAlaAspProCysLysLeuLeuAapCysG 366
 QY 2273 GCGAATTTGCCAATGTGTAAGAAGCAAGGACTCAGAGAACCGAGTGTCCCTGCAAC 2332
 Db 366 LylsPheAlaGlnCysValLysAaGlnTrpThrGluGluAlaGluCysArgCysArgG 386
 QY 2333 CAGATATGACAGCCAGCGGAGCCTGACGGTCTGGAACGAGGCTCTGTGCGCCCTGCA 2392
 Db 386 LnglYHisGlnSerHisGlyThrLeuAspTrpGlnThrLeuAsnLeuCysProProGly- 405
 QY 2393 CAAAGGAATGCGAGTCTCTCAGGAAAGGAGCTCCATGCAAGTGTCCAGATCACTCTG 2452
 Db 406 -LysThrCysValAlaGlyArgGlnAlaThrProCysArgProThrAspHisSerT 425
 QY 2453 AAATCAAGCATCAAACTAGTGTAAAAAGTTCCAAATCAACAAATACCAAGTAA 2512
 Db 425 HraGlnAlaGlnGluProGlyValLysLysLeu-----ArgGlnGlnAsnLysValV 443
 QY 2513 TCAGTAAGCAAAATCTGAATTAAGTACGCTGAGATATGAGAAATTAACCATCAAGAT 2572
 Db 443 aLysLysArgAaSerLysLeuSerAlaIleGlyPheGluGluPheGluAspGlnAspT 463
 QY 2573 GCGAAGGAAT 2583
 Db 463 IpgLgLyAen 466

 RESULT 6
 US-10-007-270-6
 ; Sequence 6, Application US/10007270
 ; Publication No. US20020160954A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hageman, Gregory S.
 ; APPLICANT: Kuehn, Markus H.
 ; APPLICANT: University of Iowa Research Foundation
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
 ; FILE REFERENCE: 020618-000120US
 ; CURRENT APPLICATION NUMBER: US/10/007,270
 ; CURRENT FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: US 09/430,195
 ; PRIOR FILING DATE: 1999-10-29
 ; PRIOR APPLICATION NUMBER: US 09/183,972
 ; PRIOR FILING DATE: 1998-10-29
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 198
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Human IPM 150 amino acid sequence, isoform C
 US-10-007-270-6

 Alignment Scores:
 Pred. No.: 1.81e-77 Length: 198
 Score: 1038.00 Matches: 197
 Percent Similarity: 99.49% Conservative: 0
 Best Local Similarity: 99.49% Mismatches: 1
 Query Match: 17.67% Indels: 0
 DB: 13 Gaps: 0

 US-10-007-270-1 (1-3330) x US-10-007-270-6 (1-198)
 QY 131 ATGTATTGGAACTAGAAGCTATTTTGTGATTTTCTCCAGTTCAAGGA 190
 Db 1 MetTyrLeuGluThrArgAlaIlePheValPheTrpIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAAGATATCTCCATTAAATACATACCTCTGGAACCTAAAGACATAGACAATCCCA 250
 Db 21 ThrLysAspIleSerIleAsnLleTyrHisSerGluThrLysAspIleAsnAspProPro 40

QY 251 AGAATGAAACATCTGAAGTACTGAAAAAATGTACAAATGTCAACTATGAGACGAATA 310
 Db |||||
 QY 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgile 60
 Db |||||
 QY 311 TTCCGATTGGCAAGCATCGAACAAAGATCGCATTTTCCCAACGGGGTTAAAGTC 370
 Db |||||
 QY 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
 Db |||||
 QY 371 TGTCACAGGAATCCATCAACAGATTTTAGACAGTCTTCAAGCTATTATATAGATTGAGA 430
 Db |||||
 QY 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLysVal 100
 Db |||||
 QY 431 GTGTGTCCAGAGCAGTATGGAGAGCATATCGAATCTTTCGGATCGCATCCCTGACACA 490
 Db |||||
 QY 101 ValCysGlnGluAlaLysArgThrLysArgSerAlaPheLeuAspArgileProAspThr 120
 Db |||||
 QY 491 GGGGAATATCAGCACTGGTCAGCATCTCCAGCAGGACCTGGATCTTCCAGCAGAGTAATAACAG 610
 Db |||||
 QY 611 AGAAGTTCCCTGACAGAAAGATGAAATATCTGCAGAGACATTTGGAGAGCCTGGT 670
 Db |||||
 QY 161 ArgSerPheProAspArgLysAspGluLeuSerAlaGluLysThrLeuGluProGly 180
 Db |||||
 QY 671 GAAACATTGTCTATTCAACAGCAATCTACATTTCAAGACTTTGGGCGAGTATTC 724
 Db |||||
 QY 181 GluThrIleValIleSerThrAlaIleLysSerLysThrTrpAlaValPhe 198
 Db |||||

RESULT 7

US-10-007-270-17
 ; Sequence 17, Application US/10007270
 ; Publication No. US20020160954A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hageman, Gregory S.
 ; APPLICANT: Kuehn, Markus H.
 ; APPLICANT: University of Iowa Research Foundation
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES
 ; FILE OF INVENTION: 020618-000120US
 ; CURRENT APPLICATION NUMBER: US/10/007,270
 ; CURRENT FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: US 09/430,195
 ; PRIOR FILING DATE: 1999-10-29
 ; PRIOR APPLICATION NUMBER: US 09/183,972
 ; PRIOR FILING DATE: 1998-10-29
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 17
 ; LENGTH: 1241
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Human IPM 200 amino acid sequence, isoform A
 US-10-007-270-17

Alignment Scores:
 Pred. No.: 4,73e-67 Length: 1241
 Score: 917.00 Matches: 318
 Percent Similarity: 39.23% Conservative: 161
 Best Local Similarity: 26.04% Mismatches: 311
 Query Match: 15.61% Indels: 431
 DB: 13 Gaps: 39

US-10-007-270-1 (1-3330) x US-10-007-270-17 (1-1241)

QY 170 ATTTTCTCCAGTTCAGGAACC---AAGATATCTCATTAACATATACCATTTCTGAA 226
 Db |||||
 QY 16 IlePheValIleValIleGluLysPheProSerLeuThrAlaGlnThrLysSer--- 34

QY 227 ACTAAAGACATAGACATCCC-----CCAAGAAATGAACA 262
 Db |||||
 QY 35 IleGluGluIleGlnGluProLysSerAlaValSerPheLeuLeuProGluLysSerThr 54
 Db |||||
 QY 263 -----ACTGAAGTACTGAAAAAATGTACAAATGTCAACTATGAGACGA 307
 Db |||||
 QY 55 AspLeuSerLeuAlaThrLysLysGlnProLeuAspArgGluThrGluArgGln 74
 Db |||||
 QY 308 ATATTCCGATTGGCAAGCATCGAACAAAGATCGCATTTTCCCAACGGGGTTAAA 367
 Db |||||
 QY 75 TrpLeu-----IleArgArgArgSerIleLeuPheProAsnGlyValLys 90
 Db |||||
 QY 368 GTCTCTCCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTATTATAGATTG 427
 Db |||||
 QY 91 IleCysProAspGluSerValAlaGluAlaValAlaAsnHisValLysTyrPheLysVal 110
 Db |||||
 QY 428 AGAGTGTGTGAGAGCAGTATGGAGCATATCGGATCTTCTGGATCGCATCCCTGCAC 487
 Db |||||
 QY 111 ArgValCysGlnGluAlaValTrpGluAlaPheArgThrPheTrpAspArgLeuProGly 130
 Db |||||
 QY 488 ACAGGGGAATATCAGGACTGGGTGAGCATCTCCAGCAGGAGACCTTTCGCTCTTTGAC 547
 Db |||||
 QY 131 ArgGluGluTyrHisTyrTrpMetAsnLeuCysGluAspGlyValThrSerIlePheGlu 150
 Db |||||
 QY 548 ATTTGAAAAAATCTTCCAGCAATTTCCAGGAGCACCTGGATCTTCCAGCAGAGATAAAA 607
 Db |||||
 QY 151 MetGlyThrAsnPheSerGluSerValGluHisArgSerLeuIleMetLysLysLeuThr 170
 Db |||||
 QY 608 CAGAGAAGTTTCCCTGACAGAAAAGATGAATATCTGCAGAGAAACATTTGGAGAGCCT 667
 Db |||||
 QY 171 Tyr-----AlaLysGluThrValSerSer-----SerGlu-Le 181
 Db |||||
 QY 668 GTTGAAACCATTTGTCATTTCAACAGCAATCTACATTTCAAGACTTGGGCGAGTATTTCAA 727
 Db |||||
 QY 181 UserSerProValProValGlyAspThrSerThr-LeuGlyAspThrThrLeuSerValP 201
 Db |||||
 QY 728 GAAAAACCTCAGAGAGCAAAATTCAGATGTTGCCAACGTCCTCAGTCTGGGCTTTCCCTC 787
 Db |||||
 QY 201 roHisProGluValAspAlaTyrGluGlyAlaSerGluSerSerLeuGlu----- 217
 Db |||||
 QY 788 TCACCTCTGATGACACCTCTCTCAATGAATTTCTCGATATACATCTCAACACACCAAGA 847
 Db |||||
 QY 218 --ArgProGluLysSerIleSerAsnGluIle---GluAsnValIleGluAlaThrL 236
 Db |||||
 QY 848 TGCTCAACAACAGAGAGAAACAGATTCGCTGTGTGGAGAGCAGAGGGTGGAGCTCA 907
 Db |||||
 QY 236 ysProAlaGlyGluGlnIleAlaGluPhe-----S 246
 Db |||||
 QY 908 GCCTCTCTCTGTGTAACACAGAGTTCAAGGCGAGCTCGCTGACTCCAGTCCCATATTT 967
 Db |||||
 QY 246 erIleHisLeuLeuGlyLysGlnTyrArgGluGluLeuGlnAspSerSerSerPheHis 266
 Db |||||
 QY 968 ACCAGAGCTAGCAGGAAAGTCCCACTTCAGATGCAAAAGATATTAAAGAAATTTCCAG 1027
 Db |||||
 QY 266 isGlnHisLeuGluGluGluPheIleSerGluValGluAsnAlaPheThrGlyLeuProG 286
 Db |||||
 QY 1028 GATTCAAAAAATCCATGTTGTAGATTTAGACCCAAAGAAAGAAAGAGTGGCTCAAGCT 1087
 Db |||||
 QY 286 lYrLysGluIleArgValLeuGluPheArgSerProLysGluAsnAsp-----Serg 304
 Db |||||
 QY 1088 CCACAGAGATCAACTTTCAGGCCATCTTTAAGAGACACAGTGCAGAGCAAAAAGCCCTG 1147
 Db |||||
 QY 304 lYValAspValTyrTyrAlaValThrPhe-----AsnGlyGluAlaIleSerAsn 321
 Db |||||
 QY 1148 CAAAGT---GACCTCTGCTCTTTGATTTCCAAAGAAATTCAGAGTGAAGAGTCTATCATG 1204
 Db |||||
 QY 321 hrThrTrpAspLeuIleSerLeuHisSerAsnLysValGluAsn-----HisG 337
 Db |||||
 QY 1205 GAACCATCGAGAGGAGCAAGCAACCCAGAAATCTATCTCACAGCTACAGACCTCAAAAGGC 1264
 Db |||||
 QY 337 lYLeuValGluLeuAspLysProThrValValTyrThrIleSerAsnPheArgSpt 357
 Db |||||
 QY 1265 TGATCAGCAAGCAGCTAGAGGAA----- 1287

```

Db      357 yrlleAlaGluThrLeuGlnGlnAsnPheLeuLeuGlyAsnSerSerLeuAsnProAspP 377
QY      1288 -----GACCAATCTTTGGATGG- 1305.
Db      377 roAspSerLeuGlnLeuIleAsnValArgGlyValLeuArgHisGlnThrGluAspLeuV 397
QY      1306 -----GGGCAATTCTAGTTCACT- 1323
Db      397 alTrpAsnThrGlnSerSerSerLeuGlnAlaThrProSerSerIleLeuAspAsnThrP 417
QY      1324 -----GATCAATTCCTGGATCACTGCCA- 1347
Db      417 heGlnAlaAlaTrpProSerAlaAspGluSerIleThrSerSerIleProProLeuAspP 437
QY      1348 -----GCCTTTGGTCTGCACACCCCAATCAGAGCTGCC- 1380
Db      437 heSerSerGlyProProSerAlaThrGlyArgGluLeuTrpSerGluSerProLeuGlyA 457
QY      1380 ----- 1380
Db      457 spLeuValSerThrHisLysLeuAlaPheProSerLysMetGlyLeuSerSerSerProG 477
QY      1381 -----ACAT 1384
Db      477 luValLeuGluValSerSerLeuThrLeuHisSerValThrProAlaValLeuGlnThrG 497
QY      1385 CTTTGTGCTTTATAACAGAGGATGCTACTTTGAGTCCAGAACTTCTCTGTTGAACCCC 1444
Db      497 lyLeuProValAlaSerGluGluArgThrSerGlySerHisLeu-----ValGluAspG 515
QY      1445 AGCTTGAGACAGTGCAGCGGACGACAG- 1489
Db      515 lyLeuAlaAsnValGluSerGluAspPheLeuSerIleAspSerLeuProSerSerS 535
QY      1490 CTGTGCTCCACCTGCTGATGGCTCTAC- 1522
Db      535 erPheThrGlnProValProLysGluThrIleProSerMetGluAspSerAspValSerL 555
QY      1523 TGTCAAGACTCCA-----CCTTTCTTTATGGCATCAAGCATCTTCT 1564
Db      555 eutrSerSerProTyrlleuThrSerSerIleProPheGlyLeuAspSerLeuThrSerL 575
QY      1565 CTCTGATCATCAAGGCACACAGATACATAGGCCACTGACACAGCATG- 1614
Db      575 ysValLysAspGlnLeuLysValSerProPheLeuProAspAlaSerMetGluLysGluL 595
QY      1615 -----CTAGTACCAGGCTCACCATCC 1636
Db      595 eulIlePheAspGlyGlyLeuGlySerGlySerGlyGlnLysValAspLeuIleThrTrpP 615
QY      1637 CCACCACTGATTATCTGCAATCAGCCAACTGGCTCTGGGAATTTCAATCCACCTGCAT 1696
Db      615 roTrpSerGluThrSerSer--GluLysSerAlaGluProLeuSerLysProTrpLeuG 634
QY      1697 CTTCAAGTACAGCCGATCAAGTGCAGGTGGCAAGATATGTCAGACACTA--GATG 1753
Db      634 luAspAspAspSerLeuLeuProAlaGluIleGluAspLysLysLeuValLeuValAspL 654
QY      1754 AATGATCTGCTGCAC-----A 1771
Db      654 ysMetAspSerThrAspGlnIleSerLysHisSerLysTyrlleuHisAspAspArgSerT 674
QY      1772 CTCCTGCCCATCTGAGGTACCA----- 1794
Db      674 hrHisPheProGlnGluGluProLeuSerGlyProAlaValProIlePheAlaAspThra 694
QY      1795 --GAGCTCGACGAATATGTTCTGTGCTCCAGATCATTTCTTGAG----- 1836
Db      694 laAlaGluSerAlaSerLeuThrLeuProLysHisIleSerGluValProGlyValAspA 714
QY      1837 -----GATACCACTCTGCTCTCAGCT- 1857

```

```

Db      714 spCysSerValThrLysAlaProLeuIleLeuThrSerValAlaIleSerAlaSerThra 734
QY      1858 -----TTACAGTATATCACACTAGTT 1879
Db      734 spLysSerAspGlnAlaAspAlaIleLeuArgGluAspMetGluGlnIleThrGluSerS 754
QY      1880 CT-----A 1882
Db      754 erAsnTyrlleuTrpPheAspSerGluValSerMetValLysProAspMetGlnThrLeuT 774
QY      1883 TGACCAATGGCCCCCAAGGCCCGAGAG----- 1908
Db      774 rpThrIleLeuProGluSerGluArgValTrpThrArgThrSerSerLeuGluLysLeuS 794
QY      1908 ----- 1908
Db      794 erArgAspIleLeuAlaSerThrProGlnSerAlaAspArgLeuTrpLeuSerValThrG 814
QY      1908 ----- 1908
Db      814 lnSerThrLysLeuProProThrThrIleSerThrLeuLeuGluAspGluValIleMetG 834
QY      1908 ----- 1908
Db      834 lyValGlnAspIleSerSerLeuGluAspArgIleGlyThrAspTyrlleuProGluG 854
QY      1908 ----- 1908
Db      854 lnValGlnGluGlnAsnGlyLysValGlySerTyrlleuMetSerThrSerValHisS 874
QY      1908 ----- 1908
Db      874 erThruMetValSerValAlaTrpProThrGluGlyGlyAspAspLeuSerTyrlleuG 894
QY      1909 -----CTGTAGTGTCTTCTAGTCTCGTGTGCTACATGCGCTTCCTCA 1954
Db      894 lnThrSerGlyAlaLeuValPhePheSerLeuArgValThrAsnMetMetPheSerG 914
QY      1955 ACGACCTGTTCAACAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACACAGC 2014
Db      914 luAspLeuPheAsnLysAsnSerLeuGluLysAlaLeuGluGlnArgPheLeuGluL 934
QY      2015 TGCTGTTTCATATCTACAGTCCAACTTTACAGATTTAAGCAACTTGAATATCTTAAC 2074
Db      934 euLeuValProTyrlleuGlnSerAsnLeuThrGlyPheGlnAsnLeuLeuLysAsnP 954
QY      2075 TCAGAACCGGAGTGTGATGTGAATAAGAAATGAAGTTTGTCTAGTCTGCGGTATA 2134
Db      954 heArgAsnGlySerIleValValAsnSerArgMetLysPheAlaAsnSerValProProA 974
QY      2135 ACCTTCAACCAAGCTGTGCAGCGGCTCTGAGAGATTTTCTGCTGTCAGCCCAACAC 2194
Db      974 snValAsnAsnAlaValTyrlleuLeuGluAspPheCysThrThrAlaTyrlleuAsnThrM 994
QY      2195 TCCATCTGGAAATAGACAGCTACTCTCTCAACATTTGAACCGCTGATCAGACAGATCCCT 2254
Db      994 etAsnLeuAlaIleAspLysTyrlleuAspValGluSerGlyAspGluAlaAsnProc 1014
QY      2255 GCAAGTTCTCTGCTGGCGGCAATTTGCCCAATGTGTAAAGAACCGACGAGCTGAGGAAG 2314
Db      1014 yslLysPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrpSerGlyGluA 1034
QY      2315 CGAGTGTCTGTCGCAACACCAAGATATGACCCAG-----GGGAGCCTCG 2359
Db      1034 lalysCysArgCysPheProGlyTyrlleuSerValGluGluArgProCysGlnSerLeuU 1054
QY      2360 ACAGTCTGGAACCGAGGCTCTGTGGCCCTGTGCAACAAGGAATGCGAGTCTCTCCAGGAA 2419
Db      1054 ysAspLeuGlnProAspPheCysLeuAsnAspGlyLys--CysAspIleMetProGlyH 1073
QY      2420 AGGAGCTCTCCATGCAAGTGTCCAGATC-----ACTCTGAATATCAGCATACA 2467
Db      1073 isGlyAlaIleCysArg-CysArgValGlyGluAsnTrpTrpTyrlleuArgGlyHisCys 1092

```

| | | | |
|---|------|---|------|
| QY | 2468 | AACTAGTGTGTTAAAGATCCAAAATCAACAAATCAACAGGTATCATTAAGAAATT | 2527 |
| Db | | ::: :::: ::: | |
| QY | 1093 | GlulupheValSerGluProValIleIleGlyLeuThrIleAlaSerValGlyLeu | 1112 |
| Db | | ::: :::: | |
| QY | 2528 | CTCAATTACTGACCGTAGAATATGAAGATTAAACCATCAAGATGGGAAGAAATTAAA | 2587 |
| Db | | Leu----- | |
| QY | 1113 | Leu----- | 1113 |
| QY | 2588 | AACTGAAAATGTACAATTATCTAGCTTCTCAAGAGAGATGATTGCTCTCTCAAG | 2647 |
| Db | | ::: :::: | |
| QY | 1114 | -----ValIlePheSerAlaIleIleTyrPhePheIleArgThrLeuGln | 1128 |
| Db | | -----ValIlePheSerAlaIleIleTyrPhePheIleArgThrLeuGln | |
| QY | 2648 | GAAATGAGACAGGCGAT-----ATTCATGGGTCTATCAAAATCCAGACAT | 2692 |
| Db | | ::: :::: | |
| QY | 1129 | AlaHisHisAspArgSerGluArgGluSerProPheSerGlySerSerArgGlnProAsp | 1148 |
| Db | | ::: :::: | |
| QY | 2693 | ACAGTCAACACT--GAGAAAT-----CAGCACACCATATTTCAAATATAGAGA | 2740 |
| Db | | ::: :::: | |
| QY | 1149 | SerLeuSerSerIleGluAsnAlaValIleTyrAsnProValTyrGluSerHisArg | 1167 |
| Db | | ::: :::: | |
| RESULT 8 | | | |
| US-10-007-270-15 | | | |
| ; Sequence 15, Application US/10007270 | | | |
| ; Publication NO. US20020160954A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Hagaman, Gregory S. | | | |
| ; APPLICANT: Kuehn, Markus H. | | | |
| ; APPLICANT: University of Iowa Research Foundation | | | |
| ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR OCULAR ABNORMALITIES | | | |
| ; FILE REFERENCE: 020618-0001200S | | | |
| ; CURRENT APPLICATION NUMBER: US/10/007,270 | | | |
| ; CURRENT FILING DATE: 2001-11-08 | | | |
| ; PRIOR APPLICATION NUMBER: US 09/430,195 | | | |
| ; PRIOR FILING DATE: 1999-10-29 | | | |
| ; PRIOR APPLICATION NUMBER: US 09/183,972 | | | |
| ; PRIOR FILING DATE: 1998-10-29 | | | |
| ; NUMBER OF SEQ ID NOS: 37 | | | |
| ; SOFTWARE: PatentIn Ver. 2.1 | | | |
| ; SEQ ID NO 15 | | | |
| ; LENGTH: 185 | | | |
| ; TYPE: PRT | | | |
| ; ORGANISM: Unknown Organism | | | |
| ; FEATURE: | | | |
| ; OTHER INFORMATION: Description of Unknown Organism: Monkey species | | | |
| ; FEATURE: | | | |
| ; OTHER INFORMATION: Monkey IPM 150 amino acid sequence (partial). | | | |
| US-10-007-270-15 | | | |
| Alignment Scores: | | | |
| Pred. No.: 4.33e-61 Length: 185 | | | |
| Score: 841.00 Matches: 166 | | | |
| Percent Similarity: 85.44% Conservative: 10 | | | |
| Best Local Similarity: 80.58% Mismatches: 8 | | | |
| Query Match: 14.31% Indels: 22 | | | |
| DB: 13 Gaps: 1 | | | |
| US-10-007-270-1 (1-3330) x US-10-007-270-15 (1-185) | | | |
| QY | 347 | TTTTTCCCAACGGGGTTAAAGTCTGTCCACAGAAATCCATAACAGATTTTAGACAGT | 406 |
| Db | | | |
| QY | 2 | PhePheProAsnGlyValIleValCysProGlnGluSerMetIleLeuAlaSer | 21 |
| Db | | | |
| QY | 407 | CTTCAAGCTTATTATAGATTGAGATGTGTGTCAGAGCAGTATCGGACAGATATCGATC | 466 |
| Db | | | |
| QY | 22 | LeuGlnAlaTyrTyrArgIleuArgValCysGlnGluAlaValTrpGluAlaTyrArgIle | 41 |
| Db | | | |
| QY | 467 | TTTCTGGATCGCATCCCTGACACAGGGGAAATATCAGGACTGGGTTCAGCATCTGCCAGAG | 526 |
| Db | | | |
| QY | 42 | PheLeuAspArgIleProAspThrGlyGluTyrGlnAspTrpValSerPheCysGlnGln | 61 |
| Db | | | |
| QY | 527 | GAGACTTCTGCTCTTTTCACATTTGGAAAAAATTCAGCAATTTCCAGAGACACTGGAT | 586 |
| Db | | | |

476 CGCATTCCTGCACACAGGGAATATCAGACTGGCTGAGCATCTCCAGCAGCAGACCTTC 535
Db 27 ArgLeuProGlyArgAspGluTyArgHisTrpMetAsnLeuCysGlnAspGlyValThr 46
536 TGCTCTTTTCACATTCGAAAAAATTCAGCAATTTCCAGAGCACCCTGGATCTTCTCCAG 595
Db 47 SerValPheGluMetGlyAlaHisPheSerGlnSerValGluHisArgAsnLeuIleMet 66
596 CAGAGAATA-----AAACAGAGAGTTTCCCTGCACAGAAAGATCAAAATATCTGCA 646
Db 67 LysLysIleuAlaTyThrArgGluAlaGluSerSerSerCysLysAspGln---SerCys 85
647 GAGAAGACATTCGGAGACCT-----GGTGAACCATTTGTCAATTTTCAACAGCAATC 697
Db 86 GlyProGluLeuSerPheProValProIleGlyGluThrSerThrLeuThrGlyAlaVal 105
698 TACATTTCAAGACTTGGCAGTATTCTAGAAAACCTTCAGAAGACAAATTCAGATG 757
Db 105 ----- 105
758 TTGCCAACGTCTCACTTGGGCGCTTTCCCT-----CTCACTCTCGATGACACC 804
Db 106 SerSerAlaSerTyProGlyLeuAlaSerGluSerSerAlaAlaSerProGlnGluSer 125
805 CTCTCTCAATGAATTTCCGATTAATACACTCAACGACACCAAGATCGCTACAAAGAAAGA 864
Db 126 IleSerAsnGluIle---GluAsnValThrGluGluProThrGlnProAlaAlaGluGln 144
865 GAAACAGAAATTCGCTGTGTTCGACGACAGGGTGGAGCTCAGCGTCTCTCTGGTAAAC 924
Db 145 IleAlaGluPhe-----SerIleGlnLeuLeuGly 154
925 CAGAGTTCAAGGCAGAGCTCGCTGACTCCAGTCCCATATTCACGAGCTAGCAGAGA 984
Db 155 LysArgTySerGluGluLeuArgAspProSerSerAlaLeuTyArgLeuValGlu 174
985 AAGTCCCACTTCAGATGCAAGAAATTTAAGAACTTCAGGATTCAAAAATCCAT 1044
Db 175 GluPheIleSerGluValGluLysAlaPheThrGlyLeuProGlyTyLysGlyIleArg 194
1045 GTTTAGGATTTAGACCAAGAAGAAAGATGGCTCAAGCTCCACAGAGATGCAACT 1104
Db 195 ValLeuGluPheArgAlaProGluGluAsnAspSerGlyIleAspValHisTyAlaVal 214
1105 ACGGCATCTTTAAGACACAGACGACGAGCAAGCAAAAGCCCTGCAGCT---GACCTCTG 1161
Db 215 Thr-----PheAsnGlyGluAlaIleSerAsnThrThrTrpAspLeuIle 229
1162 TCTTTTGGATTCACAAAATTTGAAAGTGAAGATCTATCATGGAACCATGAGCAGAGAC 1221
Db 230 SerLeuHisSerAsnLysValGluAsn-----HisGlyLeuValGluMetAsp 245
1222 AAGCAACACAGAAATCTATCTCAGAGCTACAGACCTCAAAAGCGTGATCAGCAAAAGCTA 1281
Db 246 AspLysProThrAlaValTyThrIleSerAsnPheArgAspTyIleAlaGluThrIleu 265
1282 GAGGAGAGA----- 1290
Db 265 HisGlnAsnPheLeuMetGlyAsnSerSerLeuAsnProAspProLysProLeuGlnLeu 285
1291 -----CAATCTTTGGATGTG----- 1305
Db 286 IleAsnValArgGlyValLeuLeuLeuProGlnThrGluAspIleValTrpAsnThrGlnSer 305
1306 GGCACAAATCAGTTCACTTGATGAATTCCTGGATCTACTCCAGCGCTTTGGTCTTGAC--- 1362
Db 306 SerSerLeuGlnValThr-----ThrSerSerIle***ValLeuGlnProAspLeu 322
1363 ---ACCAATATCAGCTGCCACATCTTTTGGCTGTTATACAGAGAGATGCTACTTTCAGT 1419
Db 323 ProValAlaProGluGlyArgThrSerGlySerPheIleLeuGluAspGlyLeuAlaSer 342

| | | | |
|----|------|---|------|
| QY | 1420 | CCAGAA----- | 1425 |
| DB | 343 | ThrGluGluLeuGluAspThrSerIleaspGlyLeuProSerSerProLeuIleGlnPro | 362 |
| QY | 1426 | -----CTTCCTCTCTGAA----- | 1443 |
| DB | 363 | ValProLysGluThrValProProMetGluAspSerAspThrAlaLeuLeuSerThrPro | 382 |
| QY | 1444 | CAGCTTGGAG-----ACAGTGGAC----- | 1479 |
| DB | 383 | HisLeuThrSerSerAlaIleGluAspLeuThrLysAspIleGlyThrProSerGlyLeu | 402 |
| QY | 1480 | -----CCTCAGACT | 1488 |
| DB | 403 | GluSerLeuAlaSerAsnIleSerAspGlnLeuGluValIleProTrpPheProAspThr | 422 |
| QY | 1489 | TCT----- | 1491 |
| DB | 423 | SerValGluLysAspPheIlePheGluSerGlyLeuGlySerGlySerGlyLysAspVal | 442 |
| QY | 1492 | -----TGGTCTCCACCTGCTGATGGCTCTACC----- | 1518 |
| DB | 443 | AspValIleAspTrpProTrpSerGluThrSerLeuGluLysThrThrLysProLeuSer | 462 |
| QY | 1519 | -----TCCTGTCTCAGAGCTCCA----- | 1536 |
| DB | 463 | LysSerTrpSerGluGluGlnAspAlaLeuLeuProThrGluGlyArgGluLysLeuHis | 482 |
| QY | 1536 | ----- | 1536 |
| DB | 483 | IleAspGlyArgValAspSerThrGluGlnIleIleGluSerSerGluHisArgTyrGly | 502 |
| QY | 1537 | ----- | 1545 |
| DB | 503 | AspArgProIleHisPheIleGluGlu***SerHisValArgSerThrIleProIlePhe | 522 |
| QY | 1546 | ATGGCATTCAAGC----- | 1581 |
| DB | 523 | ValGluSerAlaThrProProThrSerProIlePheSerIlyHisThrSerAspValPro | 542 |
| QY | 1582 | ACCACAGATACATGGCCACTGACACAGACATGCTAGTACCA----- | 1638 |
| DB | 543 | AspIleAspSerTyrSerLeuThrLysProProPheLeuProValThrIleAlaIlePro | 562 |
| QY | 1639 | ACCAGTGATTATTCTGCAATCAGCAACAGCTGGCTCTGGGAATTTCCATCCACCTGCATCT | 1698 |
| DB | 563 | AlaSerThrLysIleThrAspGluValLeuLysGluAspMetValHisThrGluSerSer | 582 |
| QY | 1699 | TCAGAT----- | 1737 |
| DB | 583 | SerHisLysGluLeuAspSerGluValProValSerArgProAspMetGlnProValTrp | 602 |
| QY | 1738 | ----- | 1752 |
| DB | 603 | ThrMetLeuProGluSerAspThrValTrpThrArgThrSerSerLeuGlyLysLeuSer | 622 |
| QY | 1753 | GAATGGATCTGCTCGACACTCTT----- | 1776 |
| DB | 623 | ArgAspThrLeuAlaSerThrProGluSerThrAspArgLeuTrpLeuLysAlaSerMet | 642 |
| QY | 1777 | -----GCCCCCATCTGAGGTACCA----- | 1815 |
| DB | 643 | ThrGlnSerThrGluLeuProSerThrThrHisSerThrGlnLeuGluGluValIle | 662 |
| QY | 1816 | GTC----- | 1818 |
| DB | 663 | MetAlaValGlnAspIleSerLeuGluLeuAspGlnValGlyThrAspTyrTyrGlnSer | 682 |
| QY | 1819 | ----- | 1851 |
| DB | 683 | GluLeuThrGluGluGlnHisGlyLysAlaAspSerTyrValGluMetSerThrSerVal | 702 |
| QY | 1852 | TCAGCTTTACAGTATATACCACTAGTTCTATGACCATTCGCCCCCAAGGCG----- | 1902 |

| | | | | | | |
|---|---|--|------|---|---|------|
| D | b | | 703 | ---- | -HisTyrThrGluMetProIleValAlaLeuProThrylsGlyValLeu | 719 |
| | | | | | | |
| | | | | ::: | ::: | ::: |
| Q | y | | 1903 | ----- | -CGAGACGTGTAGTTCTTCACGTCCGGTGTGCATAATG | 1944 |
| | | | | | | |
| D | b | | 720 | SerHisThrGlnThrAlaGlyAlaLeuValPheSerLeuArgValThrAsnMet | 739 | |
| | | | | | | |
| Q | y | | 1945 | GCTTCTCCAAACGACCTGTTCAAACAGAGCTCTCTGGAGTACCAGAGCTCTGGAGCAACA | 2004 | |
| | | | | | | |
| D | b | | 740 | LeuPheSerGluAspLeuPheAsnLysAanSerLeuGluTyrllysAlaLeuGluGlnArg | 759 | |
| | | | | | | |
| Q | y | | 2005 | TTCACACAGCTGTGGTTCATCTACAGATCCCAATCTTACAGATTAAAGCAACTGAA | 2064 | |
| | | | | | | |
| D | b | | 760 | PheLeuGluLeuLeuAlaProTyrLeuGlnSerAsnLeuSerGlyPheGlnAsnLeuGlu | 779 | |
| | | | | | | |
| Q | y | | 2065 | ATACTTAACCTTCAGAAACGGAGTGATTGTGAATAGCAAAATCAAGATTTGCTAACGTCT | 2124 | |
| | | | | | | |
| D | b | | 780 | IleLeuSerPheArgAsnGlySerIleValValAsnSerArgValArgPheAlaGluSer | 799 | |
| | | | | | | |
| Q | y | | 2125 | GTGCCGTATACTCACCAGCGTGTGCACGGGTCTTGGAGGATTTTCGTTCTGCTCA | 2184 | |
| | | | | | | |
| D | b | | 800 | AlaProProAsnValAsnLysAlaMetTyrArgileLeuGluAspPheCysThrThrAla | 819 | |
| | | | | | | |
| Q | y | | 2185 | GCCCCAACAACTCATCTCGAAATAGACACTACTCTCTCAATGTGAACACGCTGATCAA | 2244 | |
| | | | | | | |
| D | b | | 820 | TyrGlnThrMetAsnLeuAspIleAspIysTyrSerLeuAspValGluSerGlyAspGlu | 839 | |
| | | | | | | |
| Q | y | | 2245 | GCAGATCCCTGCAGTTCCTGGCTCGCGCGAATTGCCCAATGTGTAAAGAACGAACGG | 2304 | |
| | | | | | | |
| D | b | | 840 | AlaAsnProCysLysPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrp | 859 | |
| | | | | | | |
| Q | y | | 2305 | ACTGAGAAGCGAGGTGTCGTCGAACCCAGAGATATGACAGCCAGGGG | 2352 | |
| | | | | | | |
| D | b | | 860 | SerGlyGluAlaLysCysLysCysTyrProGlyTyrLeuSerValAspGluLeuProCys | 879 | |
| | | | | | | |
| Q | y | | 2353 | ---AGCTTGCAGGGTCTGGAAACAGGCGCTCTGTGGCCGTCGACAAAGGAATGCGAGTC | 2409 | |
| | | | | | | |
| D | b | | 880 | GlnSerLeuCysAspLeuGlnProAspPheCysLeuAsnAspGlyLys---CysaspIle | 898 | |
| | | | | | | |
| Q | y | | 2410 | CTCCAGGGAACGGGAGCTCCATCAGG | 2436 | |
| | | | | | | |
| D | b | | 899 | MetProGlyHieGlyAlaIleCysArg | 907 | |
| | | | | | | |

Alignment Scores:
 Pred. No.: 6.1e-26 Length: 432
 Score: 420.50 Matches: 128
 Percent Similarity: 45.88% Conservative: 89
 Best Local Similarity: 45.88% Mismatches: 161
 Query Match: 7.16% Indels: 95
 DB: 13 Gaps: 18

US-10-007-270-1 (1-3330) x US-10-007-270-19 (1-432)

| | | | |
|----|------|--|------|
| QY | 170 | ATTTTCTTCCAGTTCACAGAACG---AAAGATATCTCCATTAACATATACCATTCCTGAA | 226 |
| Db | 16 | IlePheValLeuIleGluGlyAspPheProSerLeuThrAlaGlnThrTyrLeuSer--- | 34 |
| QY | 227 | ACTAAAGACATAGACAAATCCC-----CCAGAAATGAACA | 262 |
| Db | 35 | IleGluGluIleGluProLysSerAlaValSerPheLeuLeuProGluGluSerThr | 54 |
| QY | 263 | -----ACTGAAGTACTGAAAAATGTGACAAATGTCAACTATATGAGACGA | 307 |
| Db | 55 | AspLeuSerLeuAlaThrLysLysLysGlnProLeuAspArgArgGluThrGluArgGln | 74 |
| QY | 308 | ATATTCGATTGGCAAGCATCGACAAAAAGATCCGCATTTTCCCAACGGGGGTAA | 367 |
| Db | 75 | TrpLeu-----IleArgArgArgSerIleLeuPheProaenglyVallys | 90 |
| QY | 368 | GTCTGTCCACAGCAATCATGAAACAGATTTTACAGAGTCITCAAGCTTATTATAGATTG | 427 |
| Db | 91 | IleCysProAspGluSerValAlaGluAlaValaAsnHisValLysTyrPheLysVal | 110 |
| QY | 428 | AGACTGTGTCCAGGAGCAGTATGGGACCATATCGGATCTTTCTGGATCGCATCCCTGCAC | 487 |
| Db | 111 | ArgValCysGlnGluAlaValTrpGluAlaPheArgThrPheTrpAsp***LeuProGly | 130 |
| QY | 488 | ACAGGGCAATATCAGAGTGGGTCCAGCATCTGCCAGCAGGAGACCTTCGTGCTCTTTGAC | 547 |
| Db | 131 | ArgGluGluTyrHisTyrTrpMetAsnLeuCysLeuGlyValThrSerIlePheGlu | 150 |
| QY | 548 | ATTGGAAAAAATCTTCAGCAATTCACAGAGCAGCTGGATCTTCTCCAGCAGAGATATA | 607 |
| Db | 151 | MetGlyThrAsnPheSerGluSerValGluHisArgSerLeuIleMetLysLysLeuThr | 170 |
| QY | 608 | CAGAGAAGTTTCCCTGCACAGAAAGACAAATATCTGCACAGAGAAGACATTTGGGAGACCT | 667 |
| Db | 171 | Tyr-----AlaLysGluThrValSerSer-----SerGlu-Ie | 181 |
| QY | 668 | GGTGAAACCATTTGCTATTTACACAGCAATCTACATTTCAAGACTTGGGCGAGTATCTAA | 727 |
| Db | 181 | uSerSerProValProValGlyAspThrSerThr-LeuGlyAspThrThrLeuSerValP | 201 |
| QY | 728 | GAAACCCCTCAGAGAGCAAAATTCAGAGATGTGCCAACGTCTCACTTGGGCTTCCCTC | 787 |
| Db | 201 | roHisProGluValAspAlaTyrGluGlyAlaSerGluSerSerLeuGlu----- | 217 |
| QY | 788 | TCACTCTGTATGACACCTCTCAATGTAATTTCTCGATAATACATCAACACACACCAAGA | 847 |
| Db | 218 | --ArgProGluGluSerLieserAsnGluIle-----GluAsnValIleGluGluAlaThrL | 236 |
| QY | 848 | TGCTTACAAACAGAAAGACAGAAACAGAAATTCGTGTGTTGGAGGAGCAGAGGTGGAGCTCA | 907 |
| Db | 236 | ysProAlaGlyGluGlnIleAlaGluPhe-----S | 246 |
| QY | 908 | GGGTCTCTCTGGTAAACACAGAAAGTTCACGGCAGAGCTCGCTGACTCCCGAGTCCCATAT | 967 |
| Db | 246 | exIleHisLeuLeuGlyLysGlnTyrArgGluGluLeuGlnAspSerSerPheHisH | 266 |
| QY | 968 | ACCAGAGCTAGCAGGAAGATCCCACTTCAGATGCAGAAAGATATTTAAGAAATCTCCAG | 1027 |
| Db | 266 | ieGlnHisLeuLeuGluGluPheIleSerGluValGluAsnAlaPheThrGlyLeuProG | 286 |
| QY | 1028 | GATTCAAAAAATCCATGTGTAGATTATGACCAAGAAAGAAAGATAGTGGCTCAGCT | 1087 |
| Db | 286 | IYTyLysGluIleArgValLeuGluPheArgSerProLysGluAsnAsp-----SerG | 304 |


```

; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-158

Alignment Scores:
Pred. No.: 1,62e-06 Length: 1255
Score: 189.50 Matches: 92
Percent Similarity: 41.48% Conservative: 54
Best Local Similarity: 26.14% Mismatches: 131
Query Match: 3.23% Indels: 75
DB: 14 Gaps: 19

US-10-007-270-1 (1-3330) x US-10-171-311-158 (1-1255)
QY 1345 CCAGCTTTGGT-----CCTGACCCCAATCAGAGTCCACATCTTTGCT 1392
Db 814 ProAlaHisGlyValThrSerAlaProAspThrArg---ProAlaProGlySerThrAla 832
QY 1393 GTTATAACAGAGGATGCTACTTTGAGTCCAGACTCTCTCTGTTGACCCCGCTTGAG 1452
Db 833 ProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly-----Ser 850
QY 1453 ACAGTCGAGGAGCAGAGCATGCTCTA-----CCTGACACT----- 1498
Db 851 ThrAlaProProAla---HisGlyValThrSerAlaProAspThrArgProAlaProGly 869
QY 1489 TCTTGCTCTCCACTGCTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1524
Db 870 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 889
QY 1525 TCAGAGCTCTCCACTCTTTCTTTATGATCAGCATCTCTCTCTCTCTCTCTCTCTCTCTCT 1575
Db 890 SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro 906
QY 1576 ---CAAGGACACACAGAT-----ACAATGGCCACTGACACAGACATG 1614
Db 907 AlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 925
QY 1615 CTAGTACAGGCTCCACATCCGACAGT-----GATTATCTGCAATCGACCAA 1665
Db 926 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 945
QY 1666 CTGGCTCTGGGAATTTTCATCTCCACT-----GCATCTCTCAGATGACGCGCATCA 1716
Db 946 ProAlaLeuGlySerThrAlaProProValHisAsnValThrSerAlaSerGlySerAla 965
QY 1717 AGTCAGGTGGCGAAGATATGCTCAGACACTAGATGAATGGATCTGCTGACACTCTCT 1776
Db 966 SerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrPro 985
QY 1777 GCCCATCTGAGGTACAGAGCTCAGCGAATATGTTCTGTCCTCAGATCTTTCTTGAG 1836
Db 986 AlaSerGlySerThrPro-----PheSerIleProSerHisSerAsp 1000
QY 1837 GATACCACTCTCTGTC-----TCA 1854

```

```

Db 1001 ThrProThrThrLeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSer 1020
QY 1855 GCTTTACAGTATATACACACTAGTTCTATGACACTTGGCCCAAGGCGCGAGAGCTGGTA 1914
Db 1021 SerValProProLeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyVal 1040
QY 1915 GTGTTCTTC-----AGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2028
Db 1041 SerPhePhePheLeuSerPheHisIleSerAsnLeuGlnPheAsnSerSerLeuGluAsp 1060
QY 1969 AAGAGCTCTCTGAGTACCGAGCTCTGAGGACCAATTCACACAGCTGCTGCTGCTGCTGCT 2028
Db 1061 ProSerThrAspThrTyrlGlnGlnLeuGlnArgPheIleSerGluMetPheLeuGlnIle 1080
QY 2029 CTACAGTCAATCTTACAGATTTAAGCACTTGAATTAATCTTAACTTCAGAAACGGAGT 2088
Db 1081 TyrLysGlnGly-----GlyPheLeuGlyLeuSerAsnIleLysPheArgProGlySer 1098
QY 2089 GTGATTGTGATAGCAAAATGAGTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2148
Db 1099 ValValGlnLeuThrLeuAlaPheArgGly-----ThrIleAsn 1113
QY 2149 GTGACGCGGCTC-----TTGAGGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2199
Db 1114 ValHisAspValGlnThrGlnPheAsnGlnTyrlLysThrGluAlaAlaSerArgTyrlAsn 1133
QY 2200 CTGGAATATAGACACTTCTCTCAACATTTGAACCA 2235
Db 1134 LeuThrIleSerAspValSerValSerValPro 1145

RESULT 13
US-10-177-293-311
; Sequence 311, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzetaj, Lasjos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0

```



```

; SEQ ID NO 311
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-311

Alignment Scores:
Pred. No.: 1,62e-06 Length: 1255
Score: 189.50 Matches: 92
Percent Similarity: 41.48% Conservativity: 54
Best Local Similarity: 26.14% Mismatches: 131
Query Match: 3.23% Indels: 75
DB: 14 Gaps: 19

US-10-007-270-1 (1-3330) x US-10-177-293-311 (1-1255)
QY 1345 CAGCCTTTGGT-----CCTGACACCAATCAGAGCTGCCACATCTTTTGGT 1392
Db 814 ProAlaHisGlyValThrSerAlaProAspThrArg---ProAlaProGlySerThrAla 832
QY 1393 GTTATAACAGAGGAGTCTACTTTGAGTCCAGAACTTCCTCTGTTGAACCCACCTTGAG 1452
Db 833 ProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly-----Ser 850
QY 1453 ACAGTGGACGAGCAGACAGTGTCTA-----CCTGACACT-----1488
Db 851 ThrAlaProProAla--HisGlyValThrSerAlaProAspThrArgProAlaProGly 869
QY 1489 TTTTGGTCTCACCCTGCTATGCGCTTCTACCTCCCTG-----1524
Db 870 SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 889
QY 1525 TCAGAACTCCACCTTCTTTATGGCATCAAGCATCTCTCTGACTGAT-----1575
Db 890 SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro 906
QY 1576 ----CAAGCCACACACAT-----ACATGGCCACTCACCACGATG 1614
Db 907 AlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 925
QY 1615 CTAGTACAGGCTCAGCATCCACCCACAGT-----GATTATCTGCAATCAGCAAA 1665
Db 926 ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArg 945
QY 1666 CTGCTCTGGGAATTCATCCACCT-----GCATCTCAGATCAGCCGATCA 1716
Db 946 ProAlaLeuGlySerThrAlaProProValHisAenValThrSerAlaSerGlySerAla 965
QY 1717 AGTCAGGTGGCGAAGATATGTCGACAGACACCTAGATGAATGGATCTCTGACACTCTCT 1776
Db 966 SerGlySerAlaSerThrLeuValHisAenGlyThrSerAlaArgAlaThrThrPro 985
QY 1777 GCCCATCTGAGTACAGAGCTCAGCAATATGTTCTCTCCAGATCATTTCTTGAG 1836
Db 986 AlaSerGlySerThrPro-----PheSerIleProSerHisHisSerAsp 1000
QY 1837 GATACCACTCTGTGCT-----TCA 1854
Db 1001 ThrProThrThrLeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSer 1020
QY 1855 GCTTTACAGTATATACACACTAGTCTTATGACCAATTCGCCCAAGGCGCAGAGCTGGTA 1914
Db 1021 SerValProProLeuThrSerSerAenHisSerThrSerProGlnLeuSerThrGlyVal 1040
QY 1915 GTGTTCTTTC-----AGTCTGCGTGTGCTTAACATGGCTTCTCCACAGCTGTTCAAC 1968
Db 1041 SerPhePheLeuSerPheHisIleSerAenLeuGlnPheAenSerSerLeuGluAsp 1060
QY 1969 AAGAGCTCTGAGGTACCGAGCTCTGAGCAACAATTACACAGCTGCTGTTCCATAT 2028
Db 1061 ProSerThrAspTyrTyrGlnLeuLeuGlnArgAspIleSerGluMetPheLeuGlnIle 1080
QY 2029 CTACGATCGAATCTTACAGGATTTAAGCACTTGAATATCTTAATCTCAGAAAGGGAGT 2088

```

```

Db 1081 TyrLysGlnGly-----GlyPheLeuGlyLeuSerAenIleLysPheArgProGlySer 1098
QY 2089 GTGATTCTGAATAGCAAAATCAAGTTTGTCTAGTCTTGCCGTATTAACCTCACAAGGCT 2148
Db 1099 ValValValGlnLeuThrLeuAlaPheArgGlnGly-----ThrIleAen 1113
QY 2149 GTGCACGGGGTC-----TTGGAGGATTTTCTGTCGACGCCCAACAACCTCCAT 2199
Db 1114 ValHisAspValGluThrGlnPheAenGlnTyrLysThrGluAlaAlaSerArgTyrAen 1133
QY 2200 CTGGAATAGACAGCTACTCTCTCAACATGACCA 2235
Db 1134 LeuThrIleSerAspValSerValSerAspValPro 1145

```

```

RESULT 14
US-10-097-340-212
; Sequence 212, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAWATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-212

```

```

Alignment Scores:
Pred. No.: 1.33e-06 Length: 515
Score: 188.50 Matches: 85
Percent Similarity: 40.73% Conservativity: 60
Best Local Similarity: 23.88% Mismatches: 146
Query Match: 3.21% Indels: 65
DB: 14 Gaps: 16

US-10-007-270-1 (1-3330) x US-10-097-340-212 (1-515)

```


Search completed: March 4, 2004, 18:59:38
Job time : 160 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 4, 2004, 18:39:06 ; Search time 54.5 seconds
(without alignments)
11754.781 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taacacgaaggtatctct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p pool -DEV-xmlp
-Q=/cgn2.1/USP70.spool_p/US10007270/runat_04032004.160717.5031/app_query.fasta_1.3527
-DB=PIR_78 -QFMT=fascan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THRESH=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10007270@cgn 1.1.84 @runat_04032004.160717.5031 -NCPV=6 -ICPU=3
-NO MAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 188.5 | 3.2 | 1344 | 1 A35175 | mucin 1 precursor, |
| 2 | 160.5 | 2.7 | 3507 | 2 T34513 | hypothetical prote |
| 3 | 157 | 2.7 | 1621 | 2 A82255 | hypothetical prote |
| 4 | 155.5 | 2.6 | 768 | 2 H81236 | DNA topoisomerase |
| 5 | 153 | 2.6 | 919 | 2 C86578 | type I/II secretion |
| 6 | 153 | 2.6 | 919 | 2 F72045 | type III secretion |
| 7 | 153 | 2.6 | 3678 | 2 S28916 | dystrophin - mouse |
| 8 | 152.5 | 2.6 | 768 | 2 A82009 | DNA topoisomerase |
| 9 | 152.5 | 2.6 | 2397 | 1 A55535 | versican precursor |
| 10 | 150.5 | 2.6 | 595 | 2 G64529 | outer membrane pro |
| 11 | 149.5 | 2.5 | 631 | 2 I52257 | episialin - mouse |
| 12 | 148.5 | 2.5 | 1634 | 2 T26517 | hypothetical prote |
| 13 | 148 | 2.5 | 589 | 2 S74668 | hypothetical prote |
| 14 | 147 | 2.5 | 1367 | 1 S48478 | glucan 1,4-alpha-g |

| | | | | | |
|----|-------|-----|------|----------|---------------------|
| 15 | 147 | 2.5 | 1620 | 2 S61535 | nucleotide-binding |
| 16 | 147 | 2.5 | 1751 | 2 A45604 | major blood-stage |
| 17 | 146.5 | 2.5 | 2409 | 1 A60979 | versican precursor |
| 18 | 146 | 2.5 | 1280 | 2 T00365 | hypothetical prote |
| 19 | 146 | 2.5 | 1285 | 2 H85041 | hypothetical prote |
| 20 | 145.5 | 2.5 | 809 | 2 T39626 | hypothetical prote |
| 21 | 145 | 2.5 | 630 | 2 A39344 | tumor-associated m |
| 22 | 145 | 2.5 | 1036 | 2 S73601 | protein P200 - Myc |
| 23 | 145 | 2.5 | 1864 | 2 F86378 | protein F2109.12 [|
| 24 | 145 | 2.5 | 1957 | 2 T38077 | hypothetical coile |
| 25 | 145 | 2.5 | 2484 | 2 T26216 | hypothetical prote |
| 26 | 145 | 2.5 | 2607 | 2 T26216 | hypothetical prote |
| 27 | 145 | 2.5 | 2829 | 2 A42771 | reticulocyte-bindi |
| 28 | 144 | 2.5 | 1468 | 2 A4345 | nucleoporin - rat |
| 29 | 143.5 | 2.4 | 1306 | 2 S25370 | MS2 protein - yea |
| 30 | 143 | 2.4 | 1120 | 2 J27765 | mitotic spindle as |
| 31 | 143 | 2.4 | 3225 | 2 I52300 | giantin - human |
| 32 | 143 | 2.4 | 3259 | 1 A56539 | giantin - human |
| 33 | 142.5 | 2.4 | 347 | 2 S10571 | mucin 1 precursor, |
| 34 | 142.5 | 2.4 | 772 | 2 T27907 | hypothetical prote |
| 35 | 142.5 | 2.4 | 1113 | 2 T47381 | hypothetical prote |
| 36 | 141.5 | 2.4 | 3381 | 2 T42389 | versican precursor |
| 37 | 141 | 2.4 | 2845 | 2 I49505 | adenomatous polypo |
| 38 | 141 | 2.4 | 3685 | 1 A27605 | dystrophin, muscle |
| 39 | 140.5 | 2.4 | 529 | 2 A41137 | heat shock transcr |
| 40 | 140.5 | 2.4 | 927 | 2 T09006 | scnM protein - Str |
| 41 | 140.5 | 2.4 | 1772 | 2 A45532 | major neurozoite su |
| 42 | 139.5 | 2.4 | 1928 | 2 S46773 | myosin heavy chain |
| 43 | 139.5 | 2.4 | 2357 | 2 A59249 | class VII unconven |
| 44 | 139 | 2.4 | 2145 | 2 J24747 | adenylate cyclase |
| 45 | 138.5 | 2.4 | 3187 | 2 J25837 | 364K Golgi complex |

ALIGNMENTS

RESULT 1

A35175
Nucleic 1 precursor, repetitive splice form A [validated] - human
N/Alternate names: breast carcinoma-associated Df3 antigen; core protein KP39; episialin.
ncretatic mucin; polymorphic epithelial mucin (PEM)
C/Species: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor, i
C/Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000
C/Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51
R/Ligtenberg, M.J.L.; Vos, H.L.; Gemmisen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A/Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene enc
A/Reference number: A35175; MUID:90202794; PMID:2318825
A/Accession: A35175
A/Molecule type: mRNA
A/Residues: 1-952,1033-1344 <LIG1>
A/Cross-references: GB:M32738; GB:J05288; NID:G182121; PID:AAA35804.1; PID:G182124; GB:N
A/Experimental source: splice form A
A/Note: GenBank entries HUMPEPIS1A1 and HUMPEPIS1A2 present only the amino-and carboxyl-enc
A/Accession: B35175
A/Molecule type: mRNA
A/Residues: 1-19,29-952,1033-1344 <LIG2>
A/Cross-references: GB:M32739; GB:J05288; NID:G182126; PID:AAA35806.1; PID:G182129; GB:N
A/Experimental source: splice form B
A/Note: GenBank entries HUMPEPIS1B1 and HUMPEPIS1B2 present only the amino-and carboxyl-enc
J. Biol. Chem. 265, 15286-15293, 1990
A/Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli
A/Reference number: A35886; MUID:90368715; PMID:1697589
A/Accession: A35886
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-19,29-992,1033-1344 <GEN>
A/Cross-references: GB:J05581; NID:G188869; PID:AAA59876.1; PID:G188870
A/Note: GenBank entry HUMMUCAB includes one copy of the tandemly repeated sequence
R. J. Biol. Chem. 265, 15294-15299, 1990
A/Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.

A:Reference number: A35887; MUID:90368716; PMID:2394722
A:Accession: A35887
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:G189598; PIDN:AAA60019.1; PID:G189599
A>Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated sequence than the other sequences.
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.; Eur. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein variants.
A:Reference number: S10571; MUID:90276413; PMID:2351132
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A:Cross-references: ENBL:X52229; NID:G37053; PIDN:CAA36478.1; PID:G37054
R:Abu, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated protein.
A:Reference number: A36735; MUID:90084773; PMID:2597151
A:Accession: A36735
A:Molecule type: mRNA
A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
A:Cross-references: ENBL:M31823; NID:G181542; PIDN:AAA35757.1; PID:G181543
R:Masuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992
A:Title: A novel core protein as well as a polymorphic epithelial mucin carry peanut agglutinin binding site.
A:Reference number: JX0235; MUID:93123189; PMID:1478919
A:Accession: JX0235
A:Molecule type: mRNA
A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
A:Experimental source: Gastric carcinoma cell
R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
FEBS Lett. 356, 130-136, 1994
A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine receptor.
A:Reference number: S51026; MUID:95080414; PMID:7988707
A:Accession: S51026
A:Contents: annotation
A>Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region are partial repeats. This protein is length polymorphic. Individuals may have between 21 and 125 copies of the repeat. The repeat is defined by SmaI nuclease sites.
C:Comment: Serine and threonine residues in the tandem repeat domain are extensively glycosylated.
C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C:Genetics:
A:Gene: GDB:MUC1; PUM
A:Cross-references: GDB:120705; OMIM:158340
A:Map position: 1q21-q23
A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C:Superfamily: polymorphic epithelial mucin
C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphism
F1-1344/Product: mucin 1 precursor, splice form A #status predicted <PRED>
F1-62/Region: mucin 1 amino-terminal non-repetitive
F1-23/Domin: signal sequence #link PRED #status predicted <SIGA>
F1-19,29-32/Domin: signal sequence #link PRED #status predicted <SIGB>
F1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <SIGB>
F1-19,29-1230-1344/Product: mucin 1 precursor, splice form C #status predicted <PRED>
F1-1343-1344/Region: 20-residue repeat (GSTAPPAGVTSAPDTPAP)
F1-1245-1272/Domin: mucin 1 carboxyl-terminal non-repetitive
F1-1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted <TRN>
F1-1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Alignment Scores:
Pred. No.: 0.000187
Score: 188.50
Percent Similarity: 41.19%
Best Local Similarity: 26.14%
Query Match: 3.21%
Length: 1344
Matches: 52
Conservative: 53
Mismatch: 132
Indels: 75

DB: 1 Gaps: 19
US-10-007-270-1 (1-3330) x A35175 (1-1344)
QY 1345 CCAGCTTTGGT-----CCTGACACCAATCAGAGCTGCCACATCTTTTGGT 1392
Db 903 ProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGlySerThrAla 921
QY 1393 GTTATACAGAGGATGCTACTTGGAGTCCAGACATCTCTGTTGAACCCAGCTTGAG 1452
Db 922 ProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly-----Ser 939
QY 1453 ACAGTGCAGCGAGCAGAGCATGCTCTA-----CCTGACACT-----1488
Db 940 ThrAlaProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 958
QY 1489 TCTTGTCTCCACCTGCTATGGCTCTACCTCCCTG-----1524
Db 959 SerThrAlaProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly 978
QY 1525 TCAGAGCTCCACTTTTATGTCATCAAGCATCTCTCTCACTGAT-----1575
Db 979 SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro 995
QY 1576 ---CAAGGCACACAGAT-----ACATGCCACTGACAGCAATG 1614
Db 996 AlaProGlySerThrAlaProAlaHisGlyValThrSerAlaProAspThrArg 1014
QY 1615 CTAGTACAGCGCTCCACATCCCAACAGT-----GATTATCTGCAATCAGCA 1665
Db 1015 ProAlaProGlySerThrAlaProAlaHisGlyValThrSerAlaProAspThrArg 1034
QY 1666 CTGGCTCTGGGATTCATCACTCACT-----GCATCTTCAGATCAGACCGATCA 1716
Db 1035 ProAlaLeuGlySerThrAlaProProAlaHisAsnValThrSerAlaSerGlySerAla 1054
QY 1717 AGTCAGCTGGGAGATATGTCAGACACCTAGATGAATGATGTCGTGACACTCTCT 1776
Db 1055 SerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrPro 1074
QY 1777 GCCCATCTGAGTACCAGCTCAGCGAATATGTTCTGCCAGATCATTTCTTGGAG 1836
Db 1075 AlaSerGlySerThrPro-----PheSerileProSerHisHisSerAsp 1089
QY 1837 GATACACTCTCTGTC-----1854
Db 1090 ThrProThrThrLeuAlaSerHisSerThrThrThrAspAlaSerSerThrHisHisSer 1109
QY 1855 GCATTACAGTATATCACCACCTAGTCTATGACCATTCGCCCAAGGGCGAGAGCTGGA 1914
Db 1110 ThrValProProLeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyVal 1129
QY 1915 GTGTTCTTC-----AGTCTGCGTGTGTCAACATGACCTTCCCAAGAGCTGTTCAAC 1968
Db 1130 SerPhePheLeuSerPheHisSerAsnLeuGlnPheAsnSerSerLeuGluAsp 1149
QY 1969 AAGAGCTCTCTGGAGTACCAGCTCTGAGCAACATTCACACAGCTCTGGTTCATAT 2028
Db 1150 ProSerThrAspTyrTyrGlnGlnLeuGlnArgAspLeuSerGluMetPheLeuGln 1169
QY 2029 CTACGATCCATCTTACAGATTAAGCACTTGAATATCTTAATCTTCAAGACAGGAGT 2088
Db 1170 TyrLeuGlnGly-----GlyPheLeuGlyLeuSerAsnLeuLeuPheArgProGlySer 1187
QY 2089 GTGATTGTGAATAGCAAAATGAAGTTTCTAAGTCTGTGCGCTATTAACCTCACCAAGGCT 2148
Db 1188 ValValValGlnLeuThrLeuAlaPheArgGluGly-----ThrLeuAsn 1202
QY 2149 GTGCACGGGCTC-----TTGAGAGATTTCGTCTGTGCGAGCCCAACATTCAT 2199
Db 1203 ValHisAspValGluThrGlnPheAsnGlnTyrTyrSerGluAlaAlaSerArgTyrAsn 1222
QY 2200 CTGGAATAGACAGCTACTCTCTCAACATTGAACCA 2235

```

Db      1223 LeuThrlleSerAspValSerValSerValPro 1234
RESULT 2
T34513
hypotheical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Pavello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021, CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Insertions: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Alignment Scores:
Pred. No.:      0.0165      Length:      3507
Score:          160.50      Matches:     128
Percent Similarity: 33.23%      Conservative: 83
Best Local Similarity: 20.16%      Mismatches: 230
Query Match:      2.73%      Indels:      194
DB:              2          Gaps:         22

US-10-007-270-1 (1-3330) x T34513 (1-3507)

QY      486 ACACAGGGGAATATCAGGATGGGTGAGCATCTGCCAGCAGAGACCTTCCTGCTTTG 545
Db      1969 ThrlylglyleValSerSerThrlleAlaThrlSerSer
QY      546 ACATTGGAAAACCTTCAGCAATTCACAGAGACCTGGATCTTCACAGCAGAGAAATAA 605
Db      1982 -----GluSerThrlAlaGluProHisValThrlSerleSerSerThrlThrlSer 1999
QY      606 AACAGAGAGTTCCTCCACAGAGAAAGATGAATATCTGCAGAGAGACATGGGAGAGC 665
Db      2000 ThrlyaspMetThr-----Ser 2005
QY      666 CTGGTGAACCAATTCATTCACAGCAATCTACATTTCAAGACTTGGGACATTTCT 725
Db      2006 SerlySerProGluasnValThrlMetSerSerGluSerProGluValSerSerSer 2025
QY      726 AAGAAAACCTCAGAGAGCAATTCAGATGTTGCCAAGTCTCACTGGGCTTCCC 785
Db      2026 SerlySerThrlAlaSerGluThrlThrlValSerSer----- 2038
QY      786 TCTACTCTGATGACACCTCTCTCAATGAATTCGTATATACATCTACACAGACCAA 845
Db      2039 -----ThrlProSerGluSerSerSerSerGluAla-----ProLeuThrlSerSerPr 2054
QY      846 GATGCTACACAGAGAGAGAAACAGAAATTCGCTGTG-----TTGAGGA 890
Db      2054 oAlaThrlThrlGluValleThrlGluSerVallySerVallySerThrlProlyGluGl 2074
QY      891 GCAGAGGGTGGAGCTCAGGCTCTCTCTGTGTAACCAAGATTCAAGGAGAGCTCGCTGA 950
Db      2074 uSerSerSerGluThrlleThrlVallyLeu-----Se 2084
QY      951 CTCCAGTCCCATATTTACCAGAGCTAGCAGGAAGTCC-----CAACT 995
Db      2084 rSerlySerProGluValThrlGluSerSerVallySerSerProSerThrlProSerTh 2104
QY      996 TCAGATGCAAGAAATATTTAAGAAATCTCCAGGATTCAAAAAATCCATGTTAGGATT 1055

```

```

Db      2104 rThrlSerGlnSerValThrlSerThrlValProGluThrlSerLysSerThrlValLeuSerSe 2124
QY      1056 TAGA-----CCAAAGAAAGAAAAAGATGGCTCAAGCTCCACAGAGAT 1097
Db      2124 rGluAlaProValThrlSerThrlSerProThrlGluValHisThrlSerSerGluThrllyr 2144
QY      1098 GCAACTTACGGCCATCTTTAAGAGACACAGTGCAGAGCAAAAGCCCTCGAAGTCACT 1157
Db      2144 oSerLeuSerAlaSerSerThrlThrlGlyAspThrlAsnSerThrlProSerSerSe 2164
QY      1158 CCTGTCTTTTGAATCCAAACAATTCAGAAAGTGAAGTCTATCATGACCACTGAGGA 1217
Db      2164 rLeuAla-----SerVallySerThrlSerAlaProGluGlyThrlSerAlaSe 2180
QY      1218 GGAACAGCACCAAGAAATCTATCTCACAGCTCACAGCTCAAGAGCTCAAGCAAGC 1277
Db      2180 rValAlaProVallyLeuSerSerLeuSerProAspValSerGlnProSerThrllyr 2200
QY      1278 ACTAGAGGAGAAACAATCTTTGATGCTGGGACAAATTCAGTTCATGATGAATTCCTCG 1337
Db      2200 rPheAlaThrlGluSer-----SerThrlValGlnAlaSerGluThrlSerSerGl 2217
QY      1338 A----- 1338
Db      2217 yThrlSerVallySerThrlSerGluProGluSerHisValThrllyLeuSerleThrlSe 2237
QY      1339 -----TCCTGCTCCAGCTTTGGTCTGCACACCCCAATCAGAGCTGCCAC 1382
Db      2237 rSerAsnProSerSerSerValProValThrlSerProlySerThrlProThrlValProGl 2257
QY      1383 ATCTTTTCTGCTTATTAACAGAGAGATCTACTTTGAGTCCAGACTCTCTCTGTTGAACC 1442
Db      2257 uSerThrlGluGlnProThrlSerThrlProSerGlyGlnSerLeuThrlProMetAsnSe 2277
QY      1443 CCAGCTTGACAGATGGAGGAGAGAG-----CATGTCTA-----CCTGA 1484
Db      2277 rAsnSerGluValleThrlSerGluProHisValleSerSerSerLeuSerProAs 2297
QY      1485 CACTTCTTGTCT-----CCACCTGCTATGGCCTCTACCTCCCTGTGAGAGCTCC 1535
Db      2297 vAlSerGlnSerSerThrlThrlProAsnAsnLeuSerGluSerSerThrlValGluThrlPr 2317
QY      1536 ACCTTTCTTTATGGCATCAAGCATCTCTCTCTG-----ACTGA 1574
Db      2317 o-----LysThrlSerSerGluValSerLeuAsnSerGluGluProSerThrlThrlGl 2334
QY      1575 TCAGAGCACCA-----GATACAATGGCCACTGACACAGACAATCTAGTACCAGG 1625
Db      2334 uAlaProThrlThrlLeuSerProAspIleLeuSerThrlThrlAsnAsnLeuSerGlnSe 2354
QY      1626 GCTCACCATCCCAACAGTGAATTTCTGCAATCAGCCAA----- 1665
Db      2354 rSerThrlValSerThrlGluAspArgSerGluIleSerSerGluAsnSerGluLysProTh 2374
QY      1666 -----CTGGCTCTGGGAATTTTCATCCACCTGCTCATCTTCA----- 1701
Db      2374 rSerAlaProGluLeuValThrlSerSerValThrlHisValAlaSerSerProAspVa 2394
QY      1702 -----GATCAGACCGCATCAAGTGCAGGTGGCGAAGATATGGTCAG 1742
Db      2394 lProThrlGluSerSerGluProAspAlaThrlGlySerThrlGluAsnIleProGl 2414
QY      1743 ACACCTAGATGAATGATCTCTCTGACACTCTCCGCCCA----- 1782
Db      2414 uAlaSerSerlyGlnThrlleSerSerThrlProThrlProAspThrlThrlAlaSerGl 2434
QY      1782 ----- 1782
Db      2434 uGluProThrllySerThrlSerMetSerProAspLeuSerThrlSerAsnValLeuSe 2454
QY      1782 ----- 1782
Db      2454 rGluSerSerThrlThrlProGluSerSerSerlySerSerProValSerSerThrlGluGl 2474

```

QY 1783 -----TCTGAGGTACCAGAGCTCAGCGATATGTTTC 1814
 Db 2474 yIleSerValValThrSerThrGluPheSerLysValProGlu---SerThrIleSerSe 2493
 QY 1815 TGTCCAGATCATTTCTTGGAGGATACCACTCTCTCTCAGCTTTTACAGATATACACAC 1874
 Db 2493 rValLeuGluGluAspLeuThrLysThrProSerProIleLeuGluGluThrThr 2513
 QY 1875 TAGTTCTATGACCATTTGCCCCAGGCGCAGAGCTGTGTAGTGTCTTCAGTCTGCGTGT 1934
 Db 2513 rAlaSerGluThrSerGluProLeuThrGluAspSerLeuThr---ValSerValArgII 2532
 QY 1935 TGCTAACATGCGCTTCTCAACGACCTGTTCACAGAGCTCT 1977
 Db 2532 eHieGluLeuThrThrSerSerGluAenValProLysGluSer 2546
 RESULT 3
 A82255
 hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C:Species: Vibrio cholerae
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: A82255
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: A82035, MUID:20406833; PMID:10952301
 A:Accession: A82255
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1621 <HEI>
 A:Cross-references: GB:AE004181; GB:AE003852; NID:G9655454; PIDN:AAP94159.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC0998
 A:Map position: 1
 Alignment Scores:
 Pred. No.: 0.0232 Length: 1621
 Score: 157.00 Matches: 126
 Percent Similarity: 35.88% Conservative: 109
 Best Local Similarity: 19.24% Mismatches: 236
 Query Match: 2.67% Indels: 184
 DB: 2 Gaps: 30
 US-10-007-270-1 (1-3330) x A82255 (1-1621)

QY 1000 ATGCAAAAGATATTAAAGAACTTCCA-----GGATTCAAAAATAATCCATGTG 1047
 Db 947 ProAspLeuGluLeuValAsnGluProValThrGluGluAlaPheThrGluLeuAspGlu 966
 QY 1048 TTAGGATTTAGACCAAAAGAAAAAGATGGCTCAAGCTCCACAGATGACCACTTACG 1107
 Db 967 LeuAspLeuProGluThrGluGluAspAlaLeuAlaAspAlaGluLeuGluProVal 986
 QY 1108 GCATCTTTAAGAGACACAGTGCAGAA-----GCAAAAAGCCCTCAAGT 1152
 Db 987 Ala-----GluSerGluValGluProGluLeuAspLeuAlaSerGluProAlaGlu 1003
 QY 1153 GACCTCTCTCTTTGATTCCACAAATTTGAAAGTGAGGAAGTCTATCATCGAACCATG 1212
 Db 1004 GluGluAlaPheThrGluLeuAsnLysLeuAspLeuProGluThr-----Thr 1019
 QY 1213 GAGGAGGACAAAGCAACCAATCTATCTC---ACAGCTACAGACCTCAAA----- 1260
 Db 1020 GluGluAspAlaLeuAlaAspAlaGluLeuSerAlaThrGluSerGluValGluSer 1039
 QY 1261 -----AGGCTGATCAGCAAGACACTAGAGGAGCAACAATCTTTGGATGTGGGCAATT 1314
 Db 1040 GluLeuGluLeuValSerGluProAlaAlaGluAlaPheThrGluLeuAspGluLeu 1059
 QY 1315 -----CAGTTCACTGATGAATTTGCTGATCACTGCCAGCCTTTGCTCTGACACC 1365
 Db 1060 AspLeuProGluThrGluGluAspAlaLeuAlaAspSerGluLeuGluProAlaAla 1079
 QY 1366 CAACTCAGAGCTGCCCAATCTTTTGGTGTGTTATAACA----- 1401
 Db 1080 GluSerGluValGluProGluLeuLeuValSerGluProValThrGluGluAlaPhe 1099
 QY 1402 -----GAGGATGCTACTTTGAGTCCA 1422
 Db 1100 ThrGluLeuAspGluLeuAspLeuProGluThrGluGluAspAlaLeuAlaAspAla 1119
 QY 1423 GAATCTCTCTCT-----GTTGAACCCAGCTTTGAGACAGTGGACGAGCA 1467
 Db 1120 GluLeuGluProAlaValGluSerGluValGluProGluLeuGluLeuAla----- 1136
 QY 1468 GAGCATGGTCTACCTGACACATTTGCTGCTCCACCTGCT-----ATGCGCTTACTCTCC 1521
 Db 1137 -----SerGluProAlaGluGluAlaSerThrGlu 1147
 QY 1522 CTGTCAAGAGCTCCACTTTCTTTATGGCATCAGCATCTCTCTCTGACTGATCAAGGC 1581
 Db 1148 LeuAsnGlu-----LeuAspLeuProGluThr 1157
 QY 1582 ACCACAGATACAATGGCCACTGACACAGCAATGTCTAGTACCAGGCTCACCACCCACC 1641
 Db 1158 GluGluAspAlaLeuAlaAspAlaGluLeuGlu-----ProAla 1170
 QY 1642 AGTGATTTATTCGAATCAGCAACTGGCTGGCTGGGA----- 1677
 Db 1171 AlaGluSerGluValGluSerGluLeuGluLeuAlaSerAspLeuGluGluGluPro 1190
 QY 1678 -----ATTTCACATCCACTCTTCAGATGACAGACGATCAAGT 1719
 Db 1191 PheThrGluLeuAsnGluLeuAspLeuProGluThrGluGluAspAlaLeuAlaAsp 1210
 QY 1720 CGAGTGGCGGAGATATGGTCAGACAC-----CTAGATGAATGGATCTGTCTGACACT 1773
 Db 1211 AlaGluLeuGluProAlaValGluSerGluValGluProGluLeuGluLeuAlaThrGlu 1230
 QY 1774 CTGCCCCCATCTGAGGTA---CCAGAGCTCAGCAATATGTTCTGTCGCCAGATCAATTTC 1830
 Db 1231 ProAlaGluGluGluAlaPheProGluLeuAspGlu---LeuAspLeuProGluThr 1249
 QY 1831 TTGGAGATACCATCTCTCTGCTTACAGTTTACAGTATATACCATCTAGTCTTATGACCAT 1890
 Db 1250 GluGluAspAlaLeuAlaAspAlaGluLeuGlu-----ProAlaAlaGluAlaGluVal 1267
 QY 1891 GCCCCCAAGGCGCAGAGCTGGTGTCTTCTCAGTCTGCTGTGCTTAACATGGCCTTC 1950


```

Db      1268 GluProGlu-----LeuGluLeu 1273
      |||:::
QY      1951 TCAACGACCTGTTTCAACAGAGCTCTCTGAGTACGAGCTCTGGAGCAACAATTCACA 2010
      :::::|||||
Db      1274 AlaseAspLeuGluGluys-----GluThrPheThr 1284
      :::::
QY      2011 CAGCTGCTGGTTCCATATCTACGATCCAACTTACAGGATTTAAGCAACTTGAATACTT 2070
      :::::|||||
Db      1285 GluLeu-----AspGluLeuAspLeuPro 1292
      :::::
QY      2071 AACTTCAGAACGGAGTGTGATGTGAATAGCAAAATGAAGTTTCTAGTCTGGCGG 2130
      :::::
Db      1293 GluTyThrGluGluAspAlaLeuAlaGlnLeuGluProAlaAlaGluSerGlu 1312
      :::::
QY      2131 TATACCTTCACCAAGCTGTGCAGGGGTCTTGGAGGATTTTCTGCTGCGAGCCCAA 2190
      :::::
Db      1313 ValGluProGluLeuGluLeuAlaSerValLeuGluGlu-----GluAspSerPheThrGlu 1331
      :::::
QY      2191 CAACTCCATCTGGAATACAGACTACTCT-----CTCAACATT 2229
      |||:::
Db      1332 LeuAspGluLeuAspLeuProGluTyThrGluLysAspAlaLeuAlaAspAlaGlnLeu 1351
      :::::
QY      2230 GAACAGCT-----GATCAAGCAGATCCCTGCAAGTTCTCGGCTCGCGCAATTGCC 2283
      |||:::
Db      1352 GluProAlaAlaGluSerGluValGluPro-----GluLeuGlyAspGluThr 1367
      |||:::
QY      2284 CAATGTGAAGAACGACGAGCTGAGGAGCGAGTGTCTGCAAAACGAGATATGAC 2343
      :::::
Db      1368 GluThrLeuAlaGlnGluThrGluSerAspAlaLeuValAlaAspGluAspLeuLeuAla 1387
      :::::
QY      2344 AGC---CAGGGAGCTGCGAGCTGTGGAACGAGCTCTGTGGC 2385
      |||:::
Db      1388 SerValGluSerAlaValAspGluValGlnProGluLeuGluGly 1402
      |||:::
RESULT 4
H81236
DNA topoisomerase I NMB0118 [imported] - Neisseria meningitidis (strain MC58 serogroup B
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81236
R;Jettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiugnani, V.; Pizsa, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: H81236
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-768 <TET>
A:Cross-references: GB:AE002370; GB:AE002098; NID:g7225327; PIDN:AAF40577.1; PID:g722533
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0118
C:Superfamily: DNA topoisomerase I

Alignment Scores:
Pred. No.: 0.0242 Length: 768
Score: 155.50 Matches: 140
Percent Similarity: 33.54% Conservative: 75
Best Local Similarity: 21.84% Mismatches: 243
Query Match: 2.63% Indels: 183
DB: 2 Gaps: 30

US-10-007-270-1 (1-3330) x H81236 (1-768)

QY      859 GAAGAGAAACAGAAATTCGCTGTGTGGAGGAGGAGG-----GTGGAGCTCAGC 909
      |||:::
Db      181 GluArgGluAenGluLeuArgAlaPheGluAlaGlnGluTyTrpThrValHisLeuAsp 200
      |||:::
QY      910 GTCTCTCTGTGTAACCAAGAGTTCAAGCGAGAGCTCGTGACTCCAGTCCCATATTAC 969
      :::::

```

```

Db      201 SerHisGlyArgSerLysPheThrAlaLysLeuAlaGlnTyAsnGlyAlaLysLeu 220
      |||:::
QY      970 CAGAGCTAGCAGGAAGTCCCACTTCAGATGCAAAAGATATTTAAGAAATTTCCAGGA 1029
      :::::
Db      221 GluGlnPheAspLeuProAenGluAlaAlaGlnAlaAspValLeuLysGluGluGly 240
      :::::
QY      1030 TTCAAAAAATCCATGTGTAGGATTTAGACCAAGAAAGAAAAAGATGGCTCAAGCTCC 1089
      :::::
Db      240 ----- 240
QY      1090 ACAGAGATCAACTTACGGCCATCTTTAAGAGACACAGTGCAGAGCAAAAAGCCCTGCA 1149
      |||:::
Db      241 LysGluAlaValValThrAlaLeuGluLysLys-----LysArgSerArgAsnProAla 258
      :::::
QY      1150 AGTGACCTCTGTCTTTTGTATTCCAAACAAATTGAAGTGAGGAAGTCTATCATGAACC 1209
      :::::
Db      259 AlaProPheThrThr-----SerThr 265
      |||:::
QY      1210 ATGGAGGAGGAGCAAGCAACCAAGATCTATCTCAGCTACAGACTCAAAAGGCTGATC 1269
      |||:::
Db      266 MetGlnAlaValValThrAlaLeuGlyPheThr-----ThrAsp-----ArgThrMet 282
      :::::
QY      1270 AGCAAGCAGCTAGAGGAGCAACAATCTTTGGATGTGGG----- 1308
      |||:::
Db      283 ArgThrAlaGlnGlnLeuTyGluGlyLeuAspValGlyGlnGlyAlaLeuLeu 302
      |||:::
QY      1309 ACATTCAGTTCATCATGATCAAAATTCGTGGATCATCTGCCAGCTTTGGTCTGTACACCCA 1368
      |||:::
Db      303 ThrTyMetArgThrAspSerVal-----AsnLeuAla 313
      |||:::
QY      1369 TCAGAGCTGCCCATCTTTCTGTGTATTAACAGAGGATGCTACTTTGAGTCCAGAACTT 1428
      |||:::
Db      314 AspGluAlaLeuThrGluLeuArgHisTyrlleGluAsnLyslleGlyLysGluTyLeu 333
      |||:::
QY      1429 CCTCTGTGTGAACCCAGCTTGACAGCTGGACGGAGCAGAGCATGTCTACCTACCTACACT 1488
      |||:::
Db      334 ProSerAlaAlaLysGlnTyLysThrLysSer-----LysAsnAla 347
      |||:::
QY      1489 TCTTGTCTCCACTGTATGGCTCTACCTCCCTGTCAAGAGCTCCA----- 1536
      |||:::
Db      348 GlnGluAlaHisGluAlaIleArgProThrSerValTyArgThrProGluSerValLys 367
      |||:::
QY      1537 CTTTCTTTTATGGCATCA-----AGCATCTCTCTGACTGATCAAGGCAACACAGAT 1590
      |||:::
Db      368 ProPheLeuSerAlaAspGlnPheLysLeuTyGlnMetIleTrpGlnArgThrValAla 387
      |||:::
QY      1591 ACAATGGCCACT-----GACCAGACAATGCTAGTACCGGGCTACCATCCCC 1638
      |||:::
Db      388 CysGlnMetThrProAlaLysPheAspGlnThrThrVal-----AspIleThrValGly 405
      |||:::
QY      1639 ACCAGTATATTCCTGCAATCAGCCAACTG-----GCTCTGGGAATTTTCATCCA 1689
      |||:::
Db      406 LysGlyValPheArgValThrGlyGlnValGlnThrPheAlaGlyPheLeuSerValTy 425
      |||:::
QY      1690 CTTGCATCTTCAGATGACAGCCGATCAAGTCCAGGTGGCGAAGATATGTCAGACACCTA 1749
      |||:::
Db      426 GluGluSerSerAspAspGluGlu-----GlyGluAsp----- 436
      |||:::
QY      1750 GATGAATGATCTGTCTGACACTCTCTGCCCATCTCAGGTACCAAGCTCAGCGAATAT 1809
      |||:::
Db      437 -----SerLysLysLeuProGluMetSerGlu----- 445
      |||:::
QY      1810 GTTCTGTCCAGATCATTTCTTGAGGATACCATCTCTGTCTCAGCTTTA----- 1860
      |||:::
Db      446 -----GlyAspLysLeuProValAspLysLeuTyGlyGlu 457
      |||:::
QY      1861 CAGTATATCACCATCTAGTCTCTATGACCATTCGCCCAAGGGCCGAGAGCTGGTAGTCTTC 1920
      |||:::
Db      458 GlnHisPheThrThrPro-----ProProArgTyAsnGluAlaThrLeuVal 473
      |||:::
QY      1921 TTCAGTCTGCGTGTGTGTACATG-----GCCTTCTCCACAGACCTGTTTCAAC 1968
      |||:::

```

[illegible]

RESULT 5
 C85578
 YopC/Gen secretion protein D [imported] - Chlamydophila pneumoniae (strain J138)
 C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: C85578
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, K.; Shi
 Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A86491; PMID:20330349; PMID:10871362
 A:Accession: C85578
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-919 <STO>
 A:Cross-references: GB:BA000008; NID:g8979074; PIDN:BA098909.1; GSPDB:GN00142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: yscC

| | | |
|---------------------|--------|------------------|
| Alignment Scores: | | |
| Pred. No.: | 0.037 | Length: 919 |
| Score: | 153.00 | Matches: 133 |
| Percent Similarity: | 34.79% | Conservative: 98 |

| Best Local Similarity: | 20.03% | Mismatches: | 246 |
|---|--------|---|------|
| Query Match: | 2.60% | Indels: | 187 |
| DB: | 2 | Gaps: | 27 |
| US-10-007-270-1 (1-3330) x C86578 (1-919) | | | |
| QY | 438 | AGGAGCAGTATGGAAGCATATCGGATCTTTCTGGATGCATCCCTGACACAGGGGAAT | 497 |
| DB | 20 | LyLeLysLysIleGlyLeLeuSerGlyLeuPhePheLeuAspLeuValLeuLeuGlyVal | 39 |
| QY | 498 | ATCAGGACTGGGTCCAGCATCTGCCACAGAGAGACCTTCTGCCTCTTTGACATTTGAAAAA | 557 |
| DB | 40 | SerSerGlnArgProThrGluThrSerAlaAsnValLysHisAsnLeuAArgAspGluLys | 59 |
| QY | 558 | ACTTCAGCAATTCGCCAGGACACCTGGATCTTCTCCAGCAGAGAAATAAACAGAGAGTTT | 617 |
| DB | 60 | LeuAlaAlaCysProLys----- | 69 |
| QY | 618 | TCCTTCGACAGAAAGATGAATATCTGCAGAGAGACATTGGGAGAGCTGGTGAAACCA | 677 |
| DB | 70 | SerLeuSerAlaLysLysHisThrLysLysLysThrProGlySerIle----- | 86 |
| QY | 678 | TTGTCAATTCACAGCAATCTACATTTCAAAGACTTGGCAGATATCTAAGAAACCCCTC | 737 |
| DB | 87 | -----ProSerLysValPheSerLysPheAspAlaThrGlnAspLysThrPhe | 102 |
| QY | 738 | AGAAGACAAATTCAGATGTTCCAAAGTCTCACTGGGGCTTTCCTCTCACTCTCTGA | 797 |
| DB | 103 | GlnLysThrSerGlySer-----AlaPheProAla-LysPro-- | 114 |
| QY | 798 | TGACACCTCTCAATGAATTCGTAATACACTCAACGACACCAAGATGCCTTACAC | 857 |
| DB | 115 | -----ThrThrLeuLysGlu-----LeuGluLysArgLysProArgP | 128 |
| QY | 858 | AGAAAGAGAAACAGAAATTCCTGTG-----TTGGAGAGCAGAGGGT | 899 |
| DB | 128 | oGluLysArgThrThrAlaAspValLysArgSerProArgPheLeuProThrGlnLys | 148 |
| QY | 900 | GGAGCTCAGCGTCTCTCTGTGTAACAGAAAGTTCAAGCAGAGCTCGCTGACTCCCAGTC | 959 |
| DB | 148 | lGluLysProVal----- | 152 |
| QY | 960 | CCCATATTACGAGAGCTAGCAGGAAAGTCCCACTTCAGATCGAAAGATATTAAAGA | 1019 |
| DB | 153 | -----ProAlaLysLysGluLysLeuAspSerIleGlnValTrpGlu | 168 |
| QY | 1020 | ACTTCCAGGATTC-----AAAAAATCCATGTGTAGGATTTAGCCCAAGAAAGAAA | 1073 |
| DB | 168 | uLysGlnAsnTyrAlaArgArgAlaValAsnAlaIleAsnLeuSerIleLysGlnLe | 188 |
| QY | 1074 | AGATGCTCAAGCTCCACA-----GAGATGCAACTTACGGCCATCTTTAAGAG | 1121 |
| DB | 198 | uGluGluGlnThrSerThrValThrGluLysAspValGlnProLysThrGlnAlaThrP | 208 |
| QY | 1122 | ACAGTCCAGAA-----GCAAAAGCCCTCGAAGTGACCTC----- | 1158 |
| DB | 208 | oHisLysSerLysAsnValAlaSerProSerThrSerMetProGlyIleGluLysAl | 228 |
| QY | 1159 | -----CTGCTTTTGGATCCCAAAATTTAAAGTGGAGGAGCTCTATCATCGAAC | 1208 |
| DB | 228 | aAlaThrThrValAlaValProGlnAspLysSerGluGluLysVal----- | 244 |
| QY | 1209 | CATGGAGAGGACAGCAACGAGAAATCTATCTCAGACTACAGACTCAAAGAGCTGAT | 1268 |
| DB | 245 | -----LysGluLysGluLeuThrCysGluAspLeuLysAspAsnG | 262 |
| QY | 1269 | CAGCAAGCACTAGAGGAGAACATCTTTGGATGTGGGACAATTCAGTTCACATGATG | 1328 |
| DB | 262 | yThrThrValAsnPheGluAspIleSerIle-----LeuGluLeuLysGlnPheValSerLys | 281 |
| QY | 1329 | AATTCCTGGATCACTGCCAGCTTTGGTCTGTGACCCCAATCAGAGCTGCCCATCTTT | 1388 |
| DB | 281 | sIleSerGlyThrAsnPheValPhe-----AspSerAsnAspLeuGlnPheAsnVa | 298 |

| | | | |
|----|------|--|------|
| QY | 1389 | TGCTGTTATAACAGAGGATGCTACTTTTCAGTCGACGAACCTTCCTCTGTTGAAACCCGAG--- | 1446 |
| DB | 298 | lThrileValSerHieaspProThrSerValAspAspLeuSerThrIleuLeuGlnVa | 318 |
| QY | 1447 | -----CTTGACAGAGTGGAGCGGACGAGCATGTCTTA | 1479 |
| DB | 318 | lLeuLysMetHisAspLeuLysValValGluGlnGlyAsnAsnValLeuIleThrArgAs | 338 |
| QY | 1480 | -CTGACACTTCTTGGTCTCCACCTGCTATGGCTCTACTCTCCCTGTCAGAGCTCCACC | 1538 |
| DB | 338 | nProHisLeuSerLysLeuSerThrValThrAspSerSerLeuLysGlnThrCysG | 358 |
| QY | 1539 | TTCTTTTATGGCAATCAAGCATCTCTCTCTGACTGATCAAGGCCACACAGATACATGGC | 1598 |
| DB | 358 | uAlaValValThrArgValPheArgLeu | 368 |
| QY | 1599 | CACTGACCAGACAATGCTAGTACACGGGCTCACCTCCACCGAGTGATTATTCGCAAT | 1658 |
| DB | 369 | -----TySerValSe | 372 |
| QY | 1659 | CAGCCAACTGGCTCTGGGAATTCACATCCACCTGCATCTTCAGATGACAGCGGATCAAG | 1718 |
| DB | 372 | rProSerAlaAlaValAsnIlelleGlnProLeuLeuSerHisasp | 389 |
| QY | 1719 | TCGAGGTGGCGAGATATGGTCGACACCTAGATGAAATGGATCTG | 1764 |
| DB | 389 | eValSerAlaSerGluAlaThrArgHisValIlelleSerAspIleAlaGlyAsnValAs | 409 |
| QY | 1765 | -----TCTGACACTCTCGCCCATCTGAGGTACGACGAGCTCAGCGATATGTTCTGT | 1817 |
| DB | 409 | pLysValSerAspLeuLeuAlaLeuAspCysProGlyThrSer | 427 |
| QY | 1818 | CCGAGATCACTTTTGGAGGATACCACTCTCTCTCAGCTTTACAGTATATC | 1869 |
| DB | 427 | tThrGluTyrGluValLysTyrAlaAsnProAlaLeuValSerTyrCysGlnAspVa | 447 |
| QY | 1870 | -----ACCAGTAGTCTATGACCAATTCAGCACTGCCCCCAAGGGCCGAGA | 1907 |
| DB | 447 | lLeuGlyThrLeuAlaGluAspAspAlaPheGlnMetPheIleGlnProGlyThrAsnL | 467 |
| QY | 1908 | GCTGTAGTGTCTTCAGTCTGGGTGTGCTCAACATGGCCTTCTCCAGGACCTGTCAA | 1967 |
| DB | 467 | sIlePheValValSerSerProArgLeuAlaAsnLysAla | 480 |
| QY | 1968 | CAGAGCTCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACACAGCTGCTGTTCCATA | 2027 |
| DB | 481 | -----GluGlnLeuLysSerLeuAspValProG | 491 |
| QY | 2028 | TCTACGATCCAATCTT | 2043 |
| DB | 491 | uMetAlaHisThrLeuAspAspProAlaSerThrAlaLeuAlaLeuGlyGlyThrGlyTh | 511 |
| QY | 2044 | -ACAGGATTTAAGCAACTCAATA | 2090 |
| DB | 511 | rThrSerProLysSerLeuArgPhePheMetTyrLysLeuLysTyrGlnAsnGlyGluVa | 531 |
| QY | 2091 | GATTGTGAATACCAAAATGAGTTTGTCTAACTCTGTGCGGTATAACCTC | 2144 |
| DB | 531 | lIleAlaAsnAla | 547 |
| QY | 2145 | GGCTGTGCACGGGTCTTGGAGGATTTTGGTTCTGTGTCGCGGCCACCAACATCCATCGA | 2204 |
| DB | 547 | rAlaMetAsp | 564 |
| QY | 2205 | ANTAGACAGC | 2214 |
| DB | 564 | uValAsnAsn | 567 |

RESULT 6

F72045

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
type III secretion protein SctC CP0044 [imported] - Chlamydia pneumoniae (strains CW
1/2013

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C/Accession: F72045; G81618
R/Kalnan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A/Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.
A/Reference number: A72000; MUID:99206606; PMID:10192388
A/Accession: F72045
A/Molecule type: DNA
A/Residues: 1-919 <ARN>
A/Cross-references: GB:AE001652; GB:AE001363; NID:g4376997; PIDN:AAD18841.1; PID:g4377600
A/Experimental source: strain CML029
R/Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberger, J.F.; White, O.; Hickey,
N.; Dodson, R.; Gwinn, M.; Nelson, W.; Desoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of *Chlamydia trachomatis* Moyn and *Chlamydia pneumoniae* AR39.
A/Reference number: AB1500; MUID:20150255; PMID:10684935
A/Accession: G81618
A/Molecule type: DNA
A/Residues: 1-919 <REA>
A/Cross-references: GB:AE002168; GB:AE002161; NID:g7188982; PIDN:AAF37937.1; PID:g7188989
A/Experimental source: strain AR39, HL cells
C/Genetics:
A/Gene: yscC; CP0044

```

Alignment Scores:
Pred. No.:      0.037      Length:      919
Score:          153.00     Matches:     133
Percent Similarity: 34.79%   Conservative:  98
Best Local Similarity: 20.03% Mismatches:   246
Query Match:      2.60%     Indels:     187
DB:              2         Gaps:       27

US-10-007-270-1 (1-3330) x F72045 (11-919)

QY 438 AGGAGCAGCTATGGGAAGCTATCGGATCTTTCTGGATCGCATCCCTGACACAGGGGAAT 497
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
20 LysLysLysIleGlyLeuPhePheLeuLeuValLeuGlyVal 39

QY 498 ATCAGGACTGGGTGCAGATCTGCCAGCAGAGACCTCTCTGCTCTTTGACATTGAAAAA 557
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
40 SerSerGlnArgProThrGluThrSerAlaAsnValLysHisAsnLeuArgAspGluLys 59

QY 558 ACTTCAGCAATTCGCCAGGACGACCTGGATCTTCTCCAGCAGAGATAAAACAGAGAATT 617
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
60 LeuAlaAlaCysProLys-----AsnSerAlaAla 69

QY 618 TCCCTGCACAGAAAGATGAATATCTGCAGAGAGACATTTGGGAGAGCTGTGTAACCA 677
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
70 SerLeuSerAlaLysSerHisThrLysLysThrThrProGlySerIle----- 86

QY 678 TTGTCAATTTCAACGCAATCTACATTTTCAAGACTTGGGCAGTATTCTTAAGAAAACCTC 737
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
87 -----ProSerLysValPheSerLysPheAspAlaThrGlnAspLysThrPhe 102

QY 738 AGAAGAGCAAAATTCAGAGATGTGCCAACGTCTCACTTTGGGCCTTTTCCCTCTCACTCCCTGA 797
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
103 GlnLysThrSerGlySer-----AlaPheProAla-LysPro-- 114

QY 798 TGACACCCCTCCATCAATGAATTTCTCGATATACACTCAACGACACCAAGATGCTCAAC 857
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
115 -----ThrThrLeuLysGlu-----LeuGluGluArgLysLysProArgPzr 128

QY 858 AGAAAGAGAAAACAGAAATTCGCTGTG-----TTGGAGGAGCAGAGGGGT 899
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
128 oGluArgArgThrThrAlaAspValLysArgSerProArgPheLeuProThrGlnGluVa 148

QY 900 GGAAGCTCAGCGTCTCTCTGTGTAACACAGAGTTTCAAGGCAGAGCTCGCTACTCCCGATC 959
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
148 lGluGluProVal----- 152

QY 960 CCATATTACGAGGAGCTAGCAGGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAA 1019
Db  :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
153 -----ProAlaSerLysGluGlnLeuAspSerIleGlnValTropGlu1 168

```

QY 1968 CAAAGAGCTCTCTGAGGATCCGAGCTCTGAGGACCAACAATTCACACAGCTGTGGTTCATTA 2027
 Db ||||| ||||| |||||
 481 -----GluGlnLeuLeuLysSerLeuaspValProG 491
 QY 2028 TCTACGATCCAACTCTT----- 2043
 Db :|||
 491 uMetAlaHisThrLeuaspPProAlaSerThrAlaLeuAlaLeuGlyThrGly 511
 QY 2044 -ACAGGATTTAAGCAACTTGAATA-----CTTAACCTTCAGAAACGGGAGTGT 2090
 Db ||||| ||||| ||||| ||||| ||||| |||||
 511 rThrSerProLysSerLeuArgPhePheMetTyrLysLeuLysTyrGlnAsnGlyGluVa 531
 QY 2091 GATTCTGTAATGACAAATCAAGTTTGTAAGTCTGTGCGGTATACCTC-----ACCAA 2144
 Db ||||| :||| :||| :||| :||| :||| :||| :|||
 531 lIleAlaAsnAla-----LeuGlnAspIleGlyTyrAsnLeuTyrValThr 547
 QY 2145 GCCTGTGCACGGGCTCTTGAGGATTTCCTGCTGTCGACGCCCAACAACCTCCATCTCGA 2204
 Db ||||| :||| :||| :||| :||| :||| :||| :|||
 547 rAlaMetAsp-----GluAspPheIleAsnThrLeuAsnSerIleGlnTrpLeuG 564
 QY 2205 AATAGACAGC 2214
 Db ||||| :||| :||| :||| :||| :||| :||| :|||
 564 uValAsnAsn 567

 RESULT 7
 S28916
 dystrophin - mouse
 N:Alternate names: duchenne muscular dystrophy protein
 C:Species: Mus musculus (house mouse)
 C>Date: 22-Nov-1993 #sequence revision 03-Nov-1995 #text_change 16-Jul-1999
 C/Accession: S28916; B27162; S10922; C43837; B40134
 R:Biles, R.D.; Phelps, S.F.; Cortez, M.D.; Roberts, R.; Caskey, C.T.; Chamberlain, J.S.
 Nucleic Acids Res. 20, 1725-1731, 1992
 A>Title: Human and murine dystrophin mRNA transcripts are differentially expressed during
 A:Reference number: S28916; MUID:92253376; PMID:1579466
 A:Accession: S28916
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-3678 <BI>
 A:Cross-references: EMBL:M68859
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
 R:Koenig, M.; Hoffman, E.P.; Bertelson, C.J.; Monaco, A.P.; Feener, C.; Kunkel, L.M.
 Cell 50, 509-517, 1987
 A>Title: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary
 A:Reference number: A90897; MUID:87273512; PMID:3607877
 A:Accession: B27162
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-201 <KB>
 R:Nudel, U.; Zuk, D.; Einat, P.; Zeelon, E.; Levy, Z.; Neuman, S.; Yaffe, D.
 Nature 337, 76-78, 1989
 A>Title: Duchenne muscular dystrophy gene product is not identical in muscle and brain.
 A:Reference number: S06461; MUID:89082658; PMID:2909892
 A:Accession: S10922
 A>Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-106 <ND>
 A:Cross-references: EMBL:X14183
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A43837; MUID:9216332; PMID:1377655
 A:Accession: C43837
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P.M.; Van Ommen, G.J.; F.
 Differentiation 49, 187-193, 1992
 A>Title: Characterization and cell type distribution of a novel, major transcript of the
 A:Reference number: A40134; MUID:88018015; PMID:3659917
 A:Accession: B40134
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-3181 <RAP>
 R:Kapoor, D.; Lederfein, D.; den Dunnen, J.T.; Grootsholten, P

A: Molecule type: mRNA

A: Residues: 300-676 'F', 678-1390 <HOF>

A: Cross-references: GB:M18025

A: Genetics:

A: Introns: 11/1

C: Note: the list of introns may be incomplete

C: Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystrophin

C: Keywords: actin binding; alternative splicing

F: 14-233/Domain: alpha-actinin actin-binding domain homology <ACT>

F: 340-449/Domain: spectrin/dystrophin repeat homology <SP1>

F: 450-558/Domain: spectrin/dystrophin repeat homology <SP2>

F: 2797-2924/Domain: spectrin/dystrophin repeat homology <SP2>

F: 3048-3085/Domain: WW repeat homology <WW1>

Alignment Scores:

| Pred. No.: | 0.0519 | Length: | 3678 |
|------------------------|--------|---------------|------|
| Score: | 153.00 | Matches: | 191 |
| Percent Similarity: | 34.32% | Conservative: | 146 |
| Best Local Similarity: | 19.45% | Mismatches: | 321 |
| Query Match: | 2.60% | Indels: | 325 |
| DB: | 2 | Gaps: | 46 |

US-10-007-270-1 (1-3330) x S28916 (1-3678)

QY 173 TTTCTCCAGTTCAAGGAACCAAGATATCTCCATTAACATATACCTTCTGAAACTAAA 232

Db 1965 PheAlaGlnPheArgLeuAsnPheAlaGlnLeHisThrLeuHisGluGluThrMet 1984

QY 233 -----GACATAGACAATCCCCCAAGAAATGAACAACACT 265

Db 1985 ValValThrThrGluAspMetProLeuAspValSerTyrValProSerThrTyrLeuThr 2004

QY 266 GAAAGTACTGAAATGTTACAAATGCTCACTATGAGACGATATTC----- 313

Db 2005 GluLeuSerHisLeuGlnAlaLeuSerGluValAspHisLeuLeuAsnThrProGlu 2024

QY 314 -----GATTTCGCAAGACATCGCAAGAAATTTAGACAGTCTTCAA 412

Db 2025 LeuCysAlaTysAspPheGluAspPheLeuGln----- 2036

QY 353 CCAAGGGGGTTAAAGTCTGTCACAGGAATCCATGAACAGATTTAGACAGTCTTCAA 412

Db 2037 -----GluGluSerLeuLysAsnLeuLysAspAsnLeuGln 2048

QY 412 ----- 412

Db 2049 GlnLeuSerGlyArgGlnAspLeuLeuHisTyrLysLysThrAlaAlaLeuGlnSerAla 2068

QY 413 ---GCTTATTATAGATTGAGAGTGTGTGAGAGAGCAGTATGGGA----- 454

Db 2069 ThrSerMetGluLysValValVal-----GlnGluAlaValAlaGlnMetAspPheGlnGly 2087

QY 455 -----GCATATCGGATCTTTCTGGATCGCATCCCT----- 484

Db 2088 GluLysLeuHisArgMetTyrLysGluArgGlnGlyArgPheAspArgSerValGluLys 2107

QY 485 -----GACACAGGGGAATATCAGACTGTGTCAGAGCAGTCTGCCAGCAG 526

Db 2108 TrpArgHisPheHisTyrAspMetLysValPheAsnGlnTrpLeuAsnGluValGluGln 2127

QY 527 GAGACCTTCTGCTCTTTGACATTTGGAATAATTCAGCAATTTCCAGGAG----- 577

Db 2128 -----PhePheLysLysThrGlnAsnProGluAsnTrpGluHisAlaLysTyr 2143

QY 578 -----CACCTGGATCTTCTCAGCAGAGATATAAACAAGAGAGTTTCCCTGACAGAAA 631

Db 2144 LysTrpTyrLeuLysGluLeuGlnAspGlyIleGlyGlnArgGln----- 2158

QY 632 GATGAATATCTGACAGAGACATTTGGGAGCGCTGGTGAACCATTTGCTATTCAACA 691

Db 2159 -----AlaValValArgThrLeuAsnAlaThrGlyGluGluIleLeuGlnSer 2175

QY 692 GC-----AATCTACATTTCAAGACTTGGCGAGTATT----- 723

Db 2176 SerLysThrAspValAsnIleLeuGlnGluLysLeuGlySerLeuArgTrpHis 2195

QY 724 -----CTAAGAAAACCTCAGAACAGCAA----- 747

Db 2196 AspileCysLysGluLeuAlaGluArgArgLysArgIleGluGlnLysAsnValLeu 2215

QY 748 -----ATTCAAGATTTGCCAACAGCA----- 768

Db 2216 SerGluPheGlnArgAspLeuAsnGluPheValLeuTrpLeuGluGluAlaAspAsnIle 2235

QY 769 TCACCTGGGCTTTCCCTCTCTCATCTCTGATGACACCTCTCTCAATGAATTTCTCGAAT 828

Db 2236 AlaIleThrProLeu-----GlyAspGluGlnLeuLysGlu----- 2248

QY 829 ACACCTCAACGACACCAAGATGCTCAACAGAAAAGAGAAAGAGAAATTCGCTGTGGAG 888

Db 2249 GlnLeuGluGlnValLysLeuLeuAlaGluLeuLeuProLeuArgGlnGlyLeuLys 2268

QY 889 GAGCAGAGGGTGGAGCTCAGCGTCTCTGTGTAAACCAAGAGTTCACAGCAGAGCTCGCT 948

Db 2269 GlnLeuAsnGluThrGlyGlyAlaValLeuValSerAlaProIleArgProGluGluGln 2288

QY 949 GACTCCCACTCCCATATATACAGGAGCTACAGAAAGTCCCACTTCAGATGCAAAAG 1008

Db 2289 AspLys-----LeuGluLysLysLeuLysGlnThrAsnLeuGlnTrpLys 2304

QY 1009 ATATTATAGAACTTCCA-----GGATTCAAAAAATCCATGTGTAGATTAGA 1059

Db 2305 ValSerArgAlaLeuProGluLysGlnGlyGluLeuValHisLeuLysAspPheArg 2324

QY 1060 CCAAGAAAGAAAAGATGGCTCAAGCTCCACAGAGATGCACTTACGGCCATCTTTAAG 1119

Db 2325 GlnLeuGluGlnGln-----LeuAspHisLeuLeuLeuTrpValSerProIle 2340

QY 1120 AGACAGCTACAGAAACCAAGCCCTGCAAGTGACCTCTGTCTTTGATTCACACAAA 1179

Db 2341 ArgAsnGlnLeuGluIleTyrAsnGlnProSerGlnAlaGlyProPheAspLysGlu 2360

QY 1180 ATTGAAGTGAAGATCTTATCATGAAACCATCGAGGAGCAGACAGCAACCAAGAAATCTAT 1239

Db 2361 Ile-----GluValThrValHisGly-----LysGln----- 2369

QY 1240 CTCACAGCTACAGCTCAAGAGCTGATCAGAAAGCA-----CTAGAGGAAGACAA 1293

Db 2370 -----AlaAspValGluArgLeuLeuSerLysGlyGlnHisLeuTyrLysGluLys 2386

QY 1294 TCTTTGGATGGGGACAATTCAGTTCAGTGAATGAATTCGTGATCATCTGCCACCTTT 1353

Db 2386 ----- 2386

QY 1354 GGTCTCTGACACCAATCAGAGCTGCCCCACATCTTTTGTGTATTAACAGAGATGCTACT 1413

Db 2387 ---ProSerThrGln----- 2390

QY 1414 TTGAGTCCAGAACTTCTCTCTGTGTAACCCAGCTTCAGACAGT-----GAC 1461

Db 2391 -----ProValLysArgLysLeuGluAspLeuArgSerGluTrpGlu 2404

QY 1462 GGACAGAGCATGGTCTA-----CCTGACACTTCTTGTCTCTCCA 1500

Db 2405 AlaValAsnHisLeuLeuArgGluLeuArgThrLysGlnProAspArgAla----- 2421

QY 1501 CCTGCTATGGCTCTTACCTCCCTGTGACAGCTCCACCTTTCTTTATGGATCAAGCATC 1560

Db 2422 ProGlyLeuSerThrThrGlyAlaSer-----AlaSerGlnThr 2434

QY 1561 TTCTCTCTGCTGATCAAGGACCAACATGATGATGATGATGATGATGATGATGATGATGAT 1620

Db 2435 ValThrLeu-----ValThrGlnSerValValThrLysGluThr-----ValIle 2449

QY 1621 CCAGGGCTCAACATCCCAAGCTGATTTCTGCAATCAGCCACTGGCTCTGGGAATT 1680

Db 2450 SerLysLeuGluMetProSerSer-----LeuLeuLeuGluVal 2462
 QY 1681 TCACATCCACTGTCATCTTCAGATGACAGCCGATCAAGTCAGGTGGCGAAGATATGTC 1740
 Db 2463 -----ProAlaLeuAlaAspPheAsnArgAlaTrpThrGluLeuThrAspTrpLeu 2479
 QY 1741 AGACACTAGATGAATGGATCTGCTGACACTCTGCCCATCTGAGGTACACAGATC 1800
 Db 2480 SerLeuLeuAspArgValIleLysSerGlnArgValMetValGlyAspLeuGluAspIle 2499
 QY 1801 AGCGAATATGTT-----TCTGTCCCATCATCTTCTCGGAGTACCATCT 1845
 Db 2500 AsnGluMetIleIleLysGlnLysAlaThrLeuGlnAsp-----LeuGluGlnArgArg 2517
 QY 1846 CCTGCTCAGCTTACAGTATATCACCACTAGTTCTATGACCATTCGCCCAAG----- 1899
 Db 2518 Pro-----GlnLeuGluGluLeuIleThrAlaAlaAsnLeuLysAsnLysThrSer 2535
 QY 1900 GCGCCGAGAGCTGTAGTTCCTTCAGTCTGCGTGTCTTAACATGCGCTTCTCCACAGC 1959
 Db 2536 AsnGlnGluAlaAargThrIleIleThrAspArgIleGlnIleGlnIleThrAsp 2555
 QY 1960 -----CTGTTCAAC-----AAGAGCTCT 1977
 Db 2556 GluValGlnGluGlnLeuGlnAsnArgArgGlnGlnLeuAsnGluMetLeuLysAspSer 2575
 QY 1978 CTGGAGTACCGAGCTCTGGAGCAACAATTACACAGCTGCTG----- 2019
 Db 2576 ThrGlnTrpLeuGluAlaLysGluGluAlaGlnValIleGlnValArgGlyLys 2595
 QY 2020 -----GTTCCATATCTACGATCCATCTTACAGA 2049
 Db 2596 LeuAspSerTrpLysGluGlyProHisThrValAspAlaIleGlnLysLysIleThrGlu 2615
 QY 2050 TTATAGCACTT---GAAATACTTAACCTCAGAAACGGAGTGTGATTGTG-----AAT 2100
 Db 2616 ThrLysGlnLeuAlaLysAspLeuArgGlnArgGlnIleSerValAspValAlaAsnAsp 2635
 QY 2101 AGCAAAATGAAGTTCTGTAAGTCTGCGCTATATACCTCACAGGCTGTGACGGGTC 2160
 Db 2636 LeuAlaLeuLysLeuLeuAspTyrSerAlaAspAspThrArgLysValHisMetIle 2655
 QY 2161 TTGGAGGATTTT-----CGTTCTGCT 2181
 Db 2656 ThrGluAsnIleAsnThrSerTrpGlyAsnIleHisLysArgValSerGlnGlnAla 2675
 QY 2182 GCAGCCCAACATCCATCTCGAATAAGACAGCTACTCTCTCAACATTAACACAGCTGAT 2241
 Db 2676 AlaLeuGluThrHisArgLeuLeuGlnGlnPheProLeuAspLeuGlu----- 2692
 QY 2242 CAAGCAGATCCCTGCAAGTCTCGCTCGCGGAGATTGGCCCAATGTGTAAAGAAC--- 2298
 Db 2693 -----LysPheLeuSerTrpIleThrGluAlaGluThrAlaAsnVal 2707
 QY 2299 -----GAACGAGCTGAGGAGCGAGTGTGCG 2325
 Db 2708 LeuGlnAspAlaSerArgLysGluLysLeuLeuGluAspSerArgGlyValArgGluLeu 2727
 QY 2326 TGCAACACAGGATATGACAGCCAGGGAGCCTGGAC-----GGTCTGGAACACAGCG 2376
 Db 2728 MetLysProTrpGlnAspLeuGlnGlyGluIleGluThrHisThrAspIleTyrHisAsn 2747
 QY 2377 CTCTGTGGCCCTGGCACAAAGGAATCGAGGTCTCTCCAGGGAAGGGAGCTCCATGACG 2436
 Db 2748 LeuAspGlnAsnGlnLysIleLeuArgSerLeuGluGlySerAspGluAlaProLeu 2767
 QY 2437 TTGCCAGATCACTCTGAAATCAAGCATACAAAATAGT-----GTTAAAGTTCCAA 2490
 Db 2768 LeuGlnArgLeuAspAsnMetAsnPheLysTrpSerGlnLeuGlnLysLysSerLeu 2787
 QY 2491 AATCACAAAAATACAGGTAATCAGTAAAGAAATTTCTGAATTAATCACTGCGCGAATAT 2550
 Db 2788 AsnIleArgSerHisLeuGluAlaSerSerAspGlnTrpLysArgLeuHisLeuSerLeu 2807

QY 2551 GAAGAA 2556
 Db 2808 GlnGlu 2809

RESULT 8

A82009
 DNA topoisomerase (EC 5.99.1.2) I NMA0156 [imported] - Neisseria meningitidis (strain Z2-
 C;Species: Neisseria meningitidis
 C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C;Accession: A82009
 R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 F.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A;Reference number: A81775; MUID:2022556; PMID:10761919
 A;Accession: A82009
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-768 <PAR>
 A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CA883470.1; PID:g737892
 A;Experimental source: serogroup A, strain Z2491
 C;Genetics:
 A;Gene: topA; NMA0156
 C;Superfamily: DNA topoisomerase I
 C;Keywords: isomerase

Alignment Scores:
 Pred. No.: 0.0382 Length: 768
 Score: 152.50 Matches: 139
 Percent Similarity: 33.54% Conservative: 76
 Best Local Similarity: 21.68% Mismatches: 243
 Query Match: 2.60% Indels: 183
 DB: 2 Gaps: 30

US-10-007-270-1 (1-3330) x A82009 (1-768)

QY 859 GAAAGAGAAACACAGAAATTCGCTGTGTGGAGGACGACGAGG-----GTGGAGCTCAGC 909
 Db 181 GluArgGluAsnGluIleArgAlaPheGluAlaArgGluTyrTrpThrValHisLeuAsp 200
 QY 910 GTCTCTCTGTGTAACACAGAAAGTTCAGAGCAGAGCTCGCTGACTCCAGTCCCATATTAC 969
 Db 201 SerHisLysGlyArgSerLysPheThrAlaLysLeuAlaGlnTyrAsnGlyAlaLysLeu 220
 QY 970 CAGGAGCTAGCAGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAACTTCCAGGA 1029
 Db 221 GluGlnPheAspLeuProAsnGluAlaAlaGlnAlaAspValLeuLysGluLeuGluGly 240
 QY 1030 TTCAAAAAAATCCATCTGTGTAGGATTTAGACCACCAAGAAAGAAAGATGGCTCAAGCTCC 1089
 Db 240 ----- 240
 QY 1090 ACAGAGATGCACTTACGCGCATCTTTAAGACACACAGTGCAGACAGCAAAAGCCCTGCA 1149
 Db 241 LysGlnAlaValThrAlaIleGluLysLys-----LysArgSerArgAsnProAla 258
 QY 1150 AGTGACCTCTCTCTTTTGTATTCACAAAAATTTGAAAGTGGAGGAGTCTATCATGTGAACC 1209
 Db 259 AlaProPheThrThr-----SerThr 265
 QY 1210 ATGGAGGAGCAGCAAGCAACCAACAGAAATCTATCTCAGCTACAGACCTCAAAAGGCTGATC 1269
 Db 266 MetGlnGlnAlaValArgLysLeuGlyPheThr---ThrAsp-----ArgThrMet 282
 QY 1270 AGCAAGCACTAGAGAAAGCAATCTTTGGATGTGGGG----- 1308
 Db 283 ArgThrAlaGlnLeuTyrGluGlyIleAspValGlnGlyAlaIleGlyLeuIle 302
 QY 1309 ACAATTCAGTTCACCTGATGAATTTGCTGGATCATCTGCCAGCCTTTGGTCTGACACCA 1368
 Db 303 ThrTyrMetArgThrAspSerVal-----AsnLeuAla 313

QY 1369 TCAGAGCTGCCACATCTTTGCTGTATATACAGAGGATGCTACTTTGAGTCCAGACTT 1428
 Db |||||
 314 AspGluAlaLeuThrGluIleArgHisTyrIleGluAsnLysIleGlyLeuTyrLeu 333
 QY 1429 CCTCTCTTGAACCCACCTTGAGACAGATGGAAGAGAGAGAGATGCTTCTACCTGACACT 1488
 Db |||||
 334 ProSerAlaAlaLysGlnTyrLysThrLysSer
 QY 1489 TCTTGGTCTCCACCTGCTATGGCTCTACCTCTCTCTGTCAGAGCTCCA
 Db ::::|
 348 GlnGluAlaHisGluAlaIleArgProThrSerValTyrArgThrProGluSerValLys 367
 QY 1537 CCTTCTTTATGGCATCA-----AGCATCTCTCTCTGACTGATCAGGACGACACAGAT 1590
 Db |||||
 368 ProPheLeuSerAlaAspGlnPheLysLeuTyrGlnMetIleThrGlnArgThrValAla 387
 QY 1591 ACAATGCCACT-----GACCAGACAATGCTAGTACCAAGGCTCACCATCCCC 1638
 Db |||||
 388 CysGlnMetThrProAlaLysPheAspGlnThrVal-----AspIleThrValGly 405
 QY 1639 ACCAGTGTATTCTGCAATCAGCCAACTG-----GCTCTGGGATTTCCATCCA 1689
 Db ::::|
 406 LysGlyValPheArgValThrGlyGlnValGlnThrPheAlaGlyPheLeuSerValTyr 425
 QY 1690 CTGTCATCTCAGATGACAGCCGATCAAGTCCAGGTGGCGAGATATGTCAGACACCTA 1749
 Db |||||
 426 GluGluSerSerAspGluGlu-----GlyGluAsp----- 436
 QY 1750 GATGAATGGATCTGCTGACATCTCTGCCCCATCTGAGTACCAAGCTCAGCCGATAT 1809
 Db ::::|
 437 -----SerLysLysLeuProGluMetSerGlu--- 445
 QY 1810 GTTCTGTCCAGATCATTTCTTGAGGATACCACTCTCTGCTCAGCTTA----- 1860
 Db |||||
 446 -----GlyAspLysLeuProValAspLysLeuTyrGlyGlu 457
 QY 1861 CAGTATATCAACCTAGTTCTATGACCAATGCCCCAAGGCGCAGAGCTGTGAGTTC 1920
 Db |||||
 458 GlnHisPheThrPro-----ProProArgTyrAsnGluAlaThrLeuVal 473
 QY 1921 TTCAGTCTGGCTGTCTGAACATG-----GCCTCTCCCAACAGCTGTTCAC 1968
 Db ::::|
 474 LysAlaLeuGluGluTyrGlyIleGlyArgProSerThrTyrAlaSerIleSerThr 493
 QY 1969 AAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACAGCTGCTGTTCCATAT 2028
 Db |||||
 494 LeuLysAspArgGluTyrValThrLeuGluGlnLys----- 505
 QY 2029 CTACGATCCATCTTACAGGATTTAAGCAACTTGAATACTTAATTCAGAAACGGGAGT 2088
 Db ::::|
 506 -----ArgPheMetProThrAspThrGlyAsp 514
 QY 2089 GTGATTGTGAATGACAAATG-----AAGTTTGTAACTGTCTGCGGTATAACCTCACC 2142
 Db |||||
 515 ---IleValAsnLysPheLeuThrGluHisPheAlaGlnTyrValAspTyrHisPheThr 533
 QY 2143 AAGCTGTGTCAGGGTCTTGGAGGATTTTCTGTCGTCACGCCCCCAACACTCCATCTG 2202
 Db ::::|
 534 AlaLysLeuGluAspGlnLeuAspGluIleAlaAspGlyLysArgGlnTyrIleProLeu 553
 QY 2203 -----GAAATAGACAGC 2214
 Db MetAspLysPheThrLysProPheIleLysGlnValGluGluLysGluGlyIleGluArg 573
 QY 2215 TACTCTCTCAACATTGAACACGCTGATCAGCAGATCCC---TGC----- 2256
 Db ::::|
 574 AlaLysPheThrThrGlnGlnLeuAspGluThrCysProLysCysGlyGluHisLysLeu 593
 QY 2257 -----AAGTTCTGCGCGCGGGAATTTGCCAATGTGTA 2292
 Db |||||
 594 GlnIleLysPheGlyLysMetGlyArgPheValAlaCysAlaGlyTyrProGluCysSer 613
 QY 2293 -----AAGAACGAACGAGCTGAGGAAGCGGAGTGTCTGCTGCAAAACGAGGATAT 2340

Db 614 TyrThrArgAsnValAsnGluThrAlaGluGluAlaAlaGluArgIleAlaLysAlaGlu 633
 QY 2341 GACAGCCAGGGAGCTGACGGTCTGGAACACGAGGCTCTGTGGCCCTGSCACAAAGGAA 2400
 Db |||||
 634 AlaGluGlnAlaGluLeuAspGlyArgGluCysProLysCysGly---Gly-ArgLeuVal 652
 QY 2401 TCCGAGTCTCTCCAGGAAAGGGAGCTCCATGACAGGTTGCCAGATCACTCTGAAATCAA 2460
 Db ::::|
 652 ITrpLysTyrSerArgThrGlySerLysPheIleGlyCysAlaAsnTyrProLysCysLys 672
 QY 2461 GCATACAAACTAGTGTATAAAGTCCCAATCAACAAATACAAATACAGAGTATCAGTAA 2520
 Db ::::|
 672 HisValGlu---ProLeuGluLys---ProLysAspThrGlyValGlnCysProGlnCysL 691
 QY 2521 AGAATCTGAATTAATCACTGACCGTAGAATATGAAGAAATTAACCATCAGATTTGGAAGG 2579
 Db |||||
 691 yLysGlyAsnLeuValGluArgLys-----SerArgTyrGlyLys 704

RESULT 9
 A55535
 N:Contains: glial hyaluronate-binding protein
 N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
 versican
 C:Species: Mus musculus (house mouse)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: A55535
 R/Ito, K.; Shinomura, T.; Zako, M.; Ujita, M.; Kimata, K.
 J. Biol. Chem. 270, 958-965, 1995
 A/Title: Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan generated
 A/Reference number: A55535; MUID:95122551; PMID:7822336
 A/Accession: A55535
 A/Status: Preliminary; translated from GB/EMBL/DDBJ
 A/Molecule type: mRNA
 A/Residues: 1-2397 <RES>
 A/Cross-references: GB:D16263; NID:G862460; PIDN:BA03796.1; PID:G862461
 C/Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG
 F/1-20/Domain: signal sequence #status predicted <SIG>
 F/21-1654/Domain: versican #status predicted <MAT>
 F/167-244/Domain: link protein repeat homology <LNK1>
 F/265-346/Domain: link protein repeat homology <LNK2>
 F/2095-2126/Domain: EGF homology <EG1>
 F/2133-2164/Domain: EGF homology <EG2>
 F/2171-2291/Domain: C-type lectin homology <LCH>
 F/2298-2354/Domain: complement factor H repeat homology <FHD>

Alignment Scores:
 Pred. No.: 0.0505 Length: 2397
 Score: 152.50 Matches: 163
 Percent Similarity: 32.31% Conservative: 110
 Best Local Similarity: 19.29% Mismatches: 321
 Query Match: 2.60% Indels: 251
 DB: 34 Gaps: 1

US-10-007-270-1 (1-3330) x A55535 (1-2397)

QY 560 TTCAGCAATCCACGAGCACCCTGGATCTTCCAGCAGAGATAAATAACAGAGAGTTTC 619
 Db |||||
 480 PheLysLysSerLysGluGlyThrGluLeuLeuGluIleThrTrpLysProGluThrTyr 499
 QY 620 CTTCAAGAAAGATGAATATCTGCGAGAGACAGATTCGGAGAGCTGGTGAACCAT 679
 Db |||||
 500 ProGluThrProAspHisValSerSer-----GlyGluProAspValPhePro 515
 QY 680 GTCAATTCACAGCAATCTACATTCAAAGACTTGGCAGCATTTCTAAGAAACCTCTCAG 739
 Db ::::|
 516 ThrLeuSerSerHisAspGlyLysThrArgTrp--SerGluSerIleThrGluSerS 535
 QY 740 AAGAGCAATTCAGATGTTGCCACAGCTTCCACTTGGGCTTCCCTCTCTCACTCTGATG 799
 Db ::::|
 535 erProAsnLeuGluAsnProValHisLysGlnProLysProValProLeuPheProGluG 555

[illegible]

Biochem. Biophys. Res. Commun. 181, 121-130, 1991

A:Title: The mouse epialin (Muc1) gene and its promoter. Rapid evolution of the repeat
A:Reference number: 152257; MUID:92068178; PMID:1958179

A:Accession: J52257

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-631 <RES>

A:Cross-references: GB:M77226; NID:g199835; PIDN:AAA9754.1; PID:g199837

A:Accession: J52210

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-631 <RES>

A:Cross-references: GB:M84683; NID:g199842; PIDN:AAA9756.1; PID:g199843

C:Genetics:

A:Gene: Muc1

A:Introns: 20/1; 454/3; 472/2; 517/1; 557/3; 607/3

Alignment Scores:

| Pred. No.: | Score: | Length: |
|------------------------|--------|---------|
| 0.0574 | 631 | |
| 149.50 | 126 | |
| 30.70% | 76 | |
| Best Local Similarity: | 219 | |
| Mismatches: | 237 | |
| Query Match: | 2 | |
| DB: | 27 | |

US-10-007-270-1 (1-3330) x 152257 (1-631)

```

QY 697 CTACATTTCAAGACTGGGAGTATCTTAAGAAACCTCAGAGAGCAAAATTCAGAT 756
Db 697 CTACATTTCAAGACTGGGAGTATCTTAAGAAACCTCAGAGAGCAAAATTCAGAT 756
QY 757 GTTGCCACACTCTCACTTGGGCTTCTCCCTCTCACTCTCTGATGACACCCCTCTCAAGAA 816
Db 757 GTTGCCACACTCTCACTTGGGCTTCTCCCTCTCACTCTCTGATGACACCCCTCTCAAGAA 816
QY 817 ATTCCTGATATCACTCAACGAC---ACCAAGATGCTTACACAGAGAGCAAAACGAA 873
Db 817 ATTCCTGATATCACTCAACGAC---ACCAAGATGCTTACACAGAGAGCAAAACGAA 873
QY 88 ValLeuSerGlyThrSerSerProAlaThrAlaProValAsnSerAlaSerSerPro 107
Db 88 ValLeuSerGlyThrSerSerProAlaThrAlaProValAsnSerAlaSerSerPro 107
QY 874 TTCCGTCTGTGGAGGAGCAGGGTGGAGCTCAGCGTCTCTCTGTTAAACGAGATTC 933
Db 874 TTCCGTCTGTGGAGGAGCAGGGTGGAGCTCAGCGTCTCTCTGTTAAACGAGATTC 933
QY 108 ValAla----- 109
Db 108 ValAla----- 109
QY 934 AAGCAGAGCTCGCTGACTCCCGATCCCATATTACAGGAGCTACAGAGAAAGTCCAA 993
Db 934 AAGCAGAGCTCGCTGACTCCCGATCCCATATTACAGGAGCTACAGAGAAAGTCCAA 993
QY 110 -----HisGlyAspThrSerPro----- 116
Db 110 -----HisGlyAspThrSerPro----- 116
QY 994 CTTGAGATGCAAGATATTTAAGAACTTCAGGATTCAAAAAATCCATGCTGTAGGA 1053
Db 994 CTTGAGATGCAAGATATTTAAGAACTTCAGGATTCAAAAAATCCATGCTGTAGGA 1053
QY 116 ----- 116
Db 116 ----- 116
QY 1054 TTTAGCAAAAGAAAGAAAGATGCTCAGCTCCACAGAGATGCAACTTACGGCCATC 1113
Db 1054 TTTAGCAAAAGAAAGAAAGATGCTCAGCTCCACAGAGATGCAACTTACGGCCATC 1113
QY 117 ---AlaThrSerLeuSerLysAspSerAsnSerSerProValVal----- 130
Db 117 ---AlaThrSerLeuSerLysAspSerAsnSerSerProValVal----- 130
QY 1114 TTTAAGAGACACAGTGCAGAGCAAGAAAGCCCTGCAAGTGACCTCTCTTTTGAATCC 1173
Db 1114 TTTAAGAGACACAGTGCAGAGCAAGAAAGCCCTGCAAGTGACCTCTCTTTTGAATCC 1173
QY 131 -----HisSerGlyThrSerSerAlaProAlaThr---ThrAlaProValAspSer 146
Db 131 -----HisSerGlyThrSerSerAlaProAlaThr---ThrAlaProValAspSer 146
QY 1174 AACAAAATGAAGTGAAGAGTCTATCATGGAACCATG----- 1212
Db 1174 AACAAAATGAAGTGAAGAGTCTATCATGGAACCATG----- 1212
QY 147 Thr-----SerSerProValValHisGlyGlyThrSerSerProAlaThrSerPro 163
Db 147 Thr-----SerSerProValValHisGlyGlyThrSerSerProAlaThrSerPro 163
QY 1213 -----GAGGAGCAAGCAACGAGAAATCTATCTACAGCTACAGACCTCAAAAGGCTG 1266
Db 1213 -----GAGGAGCAAGCAACGAGAAATCTATCTACAGCTACAGACCTCAAAAGGCTG 1266
QY 164 ProGlyAspSerThrSerProAspHisSerSerThrSerSerProAlaThrArg--- 182
Db 164 ProGlyAspSerThrSerProAspHisSerSerThrSerSerProAlaThrArg--- 182
QY 1267 ATCAGCAAGACATAGAGAGCAAAATCTTTTGGATGTGGGACAAATTCAGTTCATGAT 1326
Db 1267 ATCAGCAAGACATAGAGAGCAAAATCTTTTGGATGTGGGACAAATTCAGTTCATGAT 1326
QY 183 -----AlaProGluAspThrSer-----ThrAla 191
Db 183 -----AlaProGluAspThrSer-----ThrAla 191
QY 1327 GAAATTCGTGGTCACTGCGCGCTTTGGTCTGACACCCAACTCAGAGCTGCCACATCT 1386
Db 1327 GAAATTCGTGGTCACTGCGCGCTTTGGTCTGACACCCAACTCAGAGCTGCCACATCT 1386

```

```

Db 192 ValLeuSerGlyThr-----SerSerProAlaThrThrAlaProValAspSerThr 208
QY 1387 TTTGCTCTTATAACAGAGAGTGTATTGAGTCCAGAACTTCTCTCTGTTGAACCCAG 1446
Db 209 SerSerProValAlaHisAspThrSerSerProAlaThrSerLeuSerGlu----- 226
QY 1447 CTTGAGACAGTGGAGGAGCAGAGATGCTTACCTGACACATCTTTGCTCTCCACCTGCT 1506
Db 227 -----AspSerAlaSerSerProValAlaHisGlyGlyThrSerSerProAla 242
QY 1507 ATGGCTCTTACCTCCCTGTCAGAACTCCACCTTTCTTATGCGATCAGACATCTTCT 1566
Db 243 ThrSerProLeuArgAspSerThrSerSerProValHisSerAlaSerLeuGlnAsn 262
QY 1567 CTGACTGATCAAGGACACACAGATCAATGCGCCACTCACCAGACAACTGCTAGTACCAGG 1626
Db 263 IleLys-----ThrThrSerAspLeuAlaSer----- 271
QY 1627 CTCACCATCCCCACAGTGTATTCTGCAATCAGC-----CAACTGGCTGTG 1674
Db 272 -----ThrProAspHisAsnGlyThrSerValThrThrSerSerAlaLeu 287
QY 1675 GGAATTCACATCACCCTGCTCATCTTCAGATCAGACGCGATCAAGTGCAGGTGGCGAGAT 1734
Db 288 GlySerAlaThrSerProAspHisSerGlyThrSerThrThrThrAsnSerSerGluSer 307
QY 1735 ATGGTCAGACACCTAGATGAATGATCTGTGTGACACTCTCTGCCCCCATCTGAGGTACCA 1794
Db 308 Val-----LeuAlaThrThrProValTySerSerMetPro 319
QY 1795 GAGCTCAGCGAATATGTTCT-----GTCCGAGATCAT----- 1827
Db 320 PheSerThrThrLysValThrSerGlySerAlaIleProAspHisAsnGlySerSer 339
QY 1828 -----TCTTGGAGGATACCACT--- 1845
Db 340 ValLeuProThrSerSerValLeuGlySerAlaThrSerLeuValTyAsnThrSerAla 359
QY 1846 ----- 1851
Db 360 IleAlaThrThrProValSerAsnGlyThrGlnProSerValProSerGlnTyProVal 379
QY 1852 TCAGCTTTTACAGTATACACACTAGTTCATGACATTCGCCCCCAGGCGGCGAGCTG 1911
Db 380 SerProThrMetAlaThrSerSerHisSerThrIleAlaSerSerTyTySer 399
QY 1912 GTAGTG----- 1917
Db 400 ThrValProPheSerThrPheSerAsnSerSerProGlnLeuSerValGlyValSer 419
QY 1918 -----TTCCTCAGTCTCGGTGTTGCTTAACATGGCTTCTCCAAACACTGTTCAACAG 1971
Db 420 PhePhePheLeuSerPheTyIleGlnAsnHisProPheAsnSerSerLeuGluAspPro 439
QY 1972 AGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACACAGCTGCTGTTCCATATCTA 2031
Db 440 SerSerAsnTyTyTyGlnGluLeuLysAArgAsnIleSerGlyLeuPheLeuGlnIlePhe 459
QY 2032 CGATCCCACTTACAGGATTTAAGCACTTGAATACTTAACCTTCAGAAACGGAGTGTG 2091
Db 460 AsnGlyAspPheLeuGlyIleSerSerIleLys-----PheArgSerGlySerVal 476
QY 2092 ATTCTGAAATAGCAAAATGAGTTTGTAAAGTCTGTCGCGGTATTAACCTCAGAGGCTGTG 2151
Db 477 ValValGluSerThrValValPheArgGluGlyThr----- 488
QY 2152 CAGGGGTCTGGAGGATTTTGTCTGCTGCGAGCCCAACAACTCCATCTGGAATAGAC 2211
Db 489 ---PheSerAlaSerAspValLysSerGlnLeuIleGlnHis---LysLysGluAlaAsp 506
QY 2212 AGCTACTCTCTCAACATTGA----- 2241
Db 507 AspTyAsnLeuThrIleSerGluValLysValAsnGluMetGlnPheProProSerAla 526

```

```

QY 2242 CAGCAGAGATCCC-----TGCAAGTTC 2262
      |||::: |||
Db 527 GlnSerArgProGlyValProGlyTyrGlyValAlaLeuValLeuValCysIleLeu 546
      |||::: |||
QY 2263 CTGGCTCTGGGC-----GAATTGGCCAATGTGTAAAGACGAACGGACT 2307
      ::::| |||
Db 547 ValAlaLeuAlaIleValTyr-PheLeuAlaLeuAlaValCys----- 560
      |||::: |||
QY 2308 GAGGAAGCGGAGTGTGCTGCAACACCGAGGATATGACAGCCAGGCGGAGCTGGAC 2361
      ::::| |||
Db 561 -----GlnCysArgArglys-----SerTyrGlnLeuLeuP 571
      |||::: |||

RESULT 12
T26517
hypothetical protein Y18D10A.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26517
R:Haris, B.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z20226
A:Accession: T26517
A:Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-1634 <MWL>
A:Cross-references: EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1
A:Experimental source: clone Y18D10A
C:Genetics:
A:Gene: CESP:Y18D10A.1
A:Z-score:

```

| | |
|------------------------|--------|
| Alignment Scores: | |
| Pred. No.: | 0.0843 |
| Score: | 148.50 |
| Percent Similarity: | 33.11% |
| Best Local Similarity: | 20.04% |
| Query Match: | 2.53% |
| DB: | 2 |
| Length: | 1634 |
| Matches: | 181 |
| Conservative: | 118 |
| Mismatches: | 356 |
| Indels: | 249 |
| Gaps: | 35 |

| | | |
|-----|---|-----|
| 504 | ACTGGGTGACGATCTGCCACGAGAGACCTTCTGCTCTTTGACATTCGAAAACCTTCA | 563 |
| | : | |
| 583 | ThrMetSerIleThrGluAspSerGluProSerThrSerSerThrAlaAlaLys---- | 600 |
| | : | |
| 564 | GCAATTCCACGAGCACCTTGGATCTTCTCCACAGACGATAAACAAGAGTTCCTCG | 623 |
| | : | |
| 601 | -----ArgSerLysArgAlaGluSerAspGluGluGluInAspLeuLysLeu | 617 |
| | : | |
| 624 | ACGAAAAGATGAAATATCTGCACGAGACACATTTGGGAGAGCTGTGGAACCACTTGCA | 683 |
| | : | |
| 618 | ThrAsnLysSerPro-----GluLysProLysLysPro----- | 628 |
| | : | |
| 684 | TTTCAACAGCAATCTACATTTCAAAGCATTTGGCAGTATCTTCAAGAAAACCTCTCAGAGA | 743 |
| | : | |
| 629 | -----SerLysThrThrGluGluThrValGlyAspValLeuLysLysArgLeuArg--As | 646 |
| | : | |
| 744 | GCAATTTCCAGATGTTCGACAGCTCTCATTTGGGGCTTTTCCCTCTCACTCCTCATGACAC | 803 |
| | : | |
| 646 | pThrAlaLysThrThrAlaThrValIleHisThrProGlyPro-----ProLeuArgTh | 664 |
| | : | |
| 804 | CTCCTCTCAATGAAATTCCTC----- | 822 |
| | : | |
| 664 | xArgLysMetGluArgMetArgAlaProThrAlaValThrSerSerLysLysGluLysPr | 684 |
| | : | |
| 823 | -----GATATATACCTCAACGACACCAAGATGCTCTACACAGAGAAG | 863 |
| | : | |
| 684 | oLysAsnAlaGlySerAlaAspSerLysLeuGlu-----GluGluHi | 699 |
| | : | |
| 864 | AGAAACAGAAATTCGCTGTGTGTGGAGGACAGAGGGTGAGCTCAGCGCTCTCTGTGGTAA | 923 |
| | : | |
| 699 | edLysAspGluThrMetIleLeuGluGluThrLeuAspLeuProGlnInThrSergI | 719 |
| | : | |

| | | | |
|----|------|---|------|
| QY | 924 | CGAGAGTTCAAG-----GCAGAGCTCGTGTAGTCCAGTCCCGTCCCTATTACCA | 971 |
| Db | 719 | ngInGluProArgIleSerCysGlySerGluLeuLeuaspGluGln----- | 734 |
| QY | 972 | GGAGCTAGCAGGAAGTCCCACTTCAGATCCAGAAAGATATTATTAAGAACTTCAGGATT | 1031 |
| Db | 734 | ----- | 734 |
| QY | 1032 | CAAAAAATTCATGTGTTAGGATTTAGACCAAGAAAGAAAGATGGCTCAAGCTCCAC | 1091 |
| Db | 735 | -----PheaspAlaSerGluGluHisSerGlyThrValProSe | 747 |
| QY | 1092 | AGAGATGCAACTTACGCCCATCTTTAAAGAGACACAGTCAGAGAACAAGAGCCCTGCAG | 1151 |
| Db | 747 | ralaProGluLeuThr-----LysasnProAlaPr | 757 |
| QY | 1152 | TGACCTCTGTCITTTGATTCCACAAAATTGAAAGTGAGGAAGTCTATCATCGAACCAT | 1211 |
| Db | 757 | oProValProGluAlaSerGluAlaSerAlaGluProProLysIle-----AspI | 774 |
| QY | 1212 | GGAGAGGACACAAGCAACAAATCTATCTCACAGCTACAGACTCAAGAGCTGATCAG | 1271 |
| Db | 774 | eProGluGlnAlaThrPro---IleLeuAlaLeuAlaLeuAlaLeuProThrValSerPr | 793 |
| QY | 1272 | CAAGACACTAGAGAAAGACAATCTTTGGATGTGGGACAAATTCAGTTC-----ACTGA | 1325 |
| Db | 793 | othrAlaLeuGluProProLysAlaGlnGluasnProThrAlaGluLeuProThrThrSe | 813 |
| QY | 1326 | TCGAATTCCTGGATCACTGCCAGCCTTTGGTCTTGACACCAATCAGAGCTGCCCATC | 1385 |
| Db | 813 | rgIuIleSerGlyAgaIaProGlnAlaLeuProThr-SerSerGlnThrProThrSe | 833 |
| QY | 1386 | TTTTGCTGTTATACAGAGGATGCTACTTTGGATGCCAGAACTTCCTCCTGTTGAACCCCA | 1445 |
| Db | 833 | r-----GlySerAlaAlaProProValaspAspLe | 843 |
| QY | 1446 | GCTT---GAGACAGTGGACGGAGCA----- | 1467 |
| Db | 843 | uLeuSerGluIleLeuSerGlyAlaLysThrThrLysThrArgLysAlaAlaProProAl | 863 |
| QY | 1468 | -----GAGCATGGTCTTACTGTGACACTCTTCTGGTCTCC | 1499 |
| Db | 863 | avalGlnLysSerIleSerSerThrThrGlnAlaProProThrSerValGlnAlaPr | 883 |
| QY | 1500 | ACCTGTCTATGGCCTCACTCCCTCGTCAGAGCTCCACCTTCTTTATGGCATCAGCAT | 1559 |
| Db | 883 | opro-----ThrSerCysSerAlaAlaProPro-----ValaspAspLe | 896 |
| QY | 1560 | CTTCTCTCTGATCATCAGCCACACAGATACATAATGGCCATGACACCAAGATGCTACT | 1619 |
| Db | 896 | uLeuSerGluIleLeuSerGlyAlaLysThrThrLysThrThrLysThrThrLysMet--- | 915 |
| QY | 1620 | ACCAGGGCTCACCATCCCCACCAAGTGATATTCTGGCAATCAGCCAACTGGCTCTGGAA | 1679 |
| Db | 916 | -----ProProValaspGlnLysLysIleSerSerGluAlaProProI | 930 |
| QY | 1680 | TTCAATCCACTCATCTTCAGATGCAGCCGATCAAGTCGAGCTGGCCAGAGATATGTT | 1739 |
| Db | 930 | eSeraspSerAlaProThrSerValHisGlnGlnThrProLysSerProLysGlnIleSe | 950 |
| QY | 1740 | CAGACACTAGATGAATGGATCTGTCTGACACTCTCTGCCCATCTCAGGTACCAAGACT | 1799 |
| Db | 950 | uasnSerLysThrGlyLeuaspIleSeraspSerGluAlaspGluGluGluGluGluGlu | 970 |
| QY | 1800 | CAGCCAAATGTTCTTCGCCAGATCATTTCTTGAGGATACCACTCTCTGTCTCAGCTTT | 1859 |
| Db | 970 | uargGlyMetGluIleValGluGluGlu-----GluGluAlaProProIleSeraspSe | 988 |
| QY | 1860 | ACAGTATATCACCACTAGTTCCTATGACCATGCCCCCAAGGGCCGAGACTGTAGTGT | 1919 |
| Db | 988 | rLeuGlnAlaSerGluProSerSerThrAlaThrValLysProGluLysValAlaVala | 1008 |

QY 1282 GAGGAGACAACTCTTGATGTG-----GGCAATTCAGTTCACATGAT 1326
 Db |||||:||||| ||| :||| |||||
 QY 175 AlaGluGluSerPheAspAlaValProAlaIleAlaGluPheThr--- 193
 |||||:||||| ||| :||| |||||
 QY 1327 GAAATTCCTGGATCTCCAGCCCTTGCTCTGACACCAATCAGAGTCCACATCT 1386
 Db |||||:||||| ||| :||| |||||
 QY 194 GluValProGlu-----GluPheIleProAlaGluProValAlaGluPheMet 211
 |||||:||||| ||| :||| |||||
 QY 1387 TTCTCTTATACAGAGATGCTACTTTCAGTCCAGAACTCTCTCTGTTGAAACCCAG 1446
 Db |||||:||||| ||| :||| |||||
 QY 212 GluAlaThrValThrGluGlu-----ThrProGluIleAlaGluIleThrProGlu 228
 |||||:||||| ||| :||| |||||
 QY 1447 CTGACAGAGTGGAGGAGCAGCATGCTTACTGACACTCTCTGG-----TCTCCA 1500
 Db |||||:||||| ||| :||| |||||
 QY 229 GluGlu-----GluAsnTrpValAspAlaPro 237
 |||||:||||| ||| :||| |||||
 QY 1501 CCTGCTATGGCTCTACTCTCCCTGTCAGAGCTCCACTTCTTTATGGCATCAGCATC 1560
 Db |||||:||||| ||| :||| |||||
 QY 238 Pro-----ThrAlaGluLeuGluThrGluPro----- 246
 |||||:||||| ||| :||| |||||
 QY 1561 TTCTCTCTGATCAAGCCACCATATCAATGCGCATGACAGCAATGCTAGTA 1620
 Db |||||:||||| ||| :||| |||||
 QY 247 -----LeuAspTyrGinglyProValGlyGluProMetValGluGluThrPheSer 264
 |||||:||||| ||| :||| |||||
 QY 1621 CAGGGCTCACCATCCCTCC-----ACCAGTGATTTCTGCAATCAGCCACTGGCTCTG 1674
 Db |||||:||||| ||| :||| |||||
 QY 265 ProPheAspAlaProProThrGluThrThrGluValGluAlaAspLeuGluAsn 284
 |||||:||||| ||| :||| |||||
 QY 1675 GGAATTCACATCCACTGCTCTTCAGATGACAGCCGATCAAGTGGAGTGGCGAAGAT 1734
 Db |||||:||||| ||| :||| |||||
 QY 285 TrpValAspAlaProProGluAlaSerLeuProAspMetAspPheGlyGlyGluGlu 304
 |||||:||||| ||| :||| |||||
 QY 1735 ATGGTCACACACTAGATGAATGATCTGCTGACACTCTCTGCCCATCT----- 1785
 Db |||||:||||| ||| :||| |||||
 QY 305 -----AsnPheGluProMetAspLeuAla-----ThrGluLeuProAspIleAsnAsp 320
 |||||:||||| ||| :||| |||||
 QY 1786 -----GAGTACACAG 1797
 |||||:||||| ||| :||| |||||
 Db 321 GluGlyIleThrAsnProGlyGlnLeuAlaGluLeuSerValGluThrProGlu 340
 |||||:||||| ||| :||| |||||
 QY 1798 CTCACGCAATGTTCTCTCCAGATCAATTTCTGGAGGATACCAT 1845
 Db |||||:||||| ||| :||| |||||
 QY 341 AlaAlaGluThrGluGlnSerSerAspGlnPheLeuAlaGluLeuThr 356
 |||||:||||| ||| :||| |||||
 RESULT 14
 S48478
 Glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
 N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c
 C;Species: *Saccharomyces cerevisiae*
 C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999
 C;Accession: S48478; A26877; B26877; S27281; JC6123
 R;Rowley, K.
 submitted to the EMBL Data Library, October 1994
 A;Reference number: S48478
 A;Accession: S48478
 A;Molecule type: DNA
 A;Residues: 1-1367 <ROW>
 A;Cross-references: GB:247047; EMBL:M16164; NID:g172522; PID:AAA35014.1; PID:g172525
 R;Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
 A;Reference number: A91831; MUID:87194600; PMID:3106330
 A;Accession: A26877
 A;Molecule type: DNA
 A;Residues: 1-242 <YAM>
 A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35015.1; PID:g172526
 R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Clarios, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Sacchar*

A;Reference number: S27281; MUID:89031230; PMID:3141213
 A;Accession: S27281
 A;Molecule type: DNA
 A;Residues: 1-31 <PAR>
 A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
 R;Lambrecht, M.G.; Bauer, P.F.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A;Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohy
 A;Reference number: JC6123; MUID:196323237; PMID:8710886
 A;Accession: JC6123
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-1367 <LAM>
 A;Cross-references: GB:U30626; NID:g1304386; PIDN:AA49609.1; PID:g1304387
 C;Genetics: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
 A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
 A;Cross-references: MIPS:YIR019c; SGD:S0001458
 A;Map position: 9R
 C;Superfamily: Yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F;5-21/Domain: transmembrane #status predicted <TML>
 F;1350-1366/Domain: transmembrane #status predicted <TM2>

Alignment Scores:
 Pred. No.: 0.101 Length: 1367
 Score: 147.00 Matches: 108
 Percent Similarity: 36.09% Conservative: 58
 Best Local Similarity: 23.48% Mismatches: 194
 Query Match: 2.50% Indels: 100
 DB: 1 Gaps: 18

US-10-007-270-1 (1-3330) x S48478 (1-1367)

QY 830 CACTCAACGACCAAGATGCTTACACAGAAAGAGAGAAAGATTCGTGTGTGAGG 889
 |||||:||||| ||| :||| |||||
 Db 290 HisAspThrThrPro---CysThrLysLysThrThrSerLysThrCysThrLys 308
 |||||:||||| ||| :||| |||||
 QY 889 ----- 889
 Db 309 LysThrThrThrProValProThrProSerSerThrThrGluSerSerSerAlaPro 328
 |||||:||||| ||| :||| |||||
 QY 890 -----AGCAGAGGTGGAGTCTCAGCTCTCTCTGTTAAACAGAGAGTCA 934
 |||||:||||| ||| :||| |||||
 Db 329 ValProThrProSerSerThrThrGluSerSerSerAlaProValThrSerThr 348
 |||||:||||| ||| :||| |||||
 QY 935 AGCAGAGCTCGTGAATCCAGTCCCATATTTACAGAGCTAGCAGAGAGTCCACAC 994
 |||||:||||| ||| :||| |||||
 Db 349 ThrGluSerSerSerAlaProValPro-----ThrProSer-SerSerThrThrGluSe 366
 |||||:||||| ||| :||| |||||
 QY 995 TTCAGATCCAAAGATATTAAAGAACTTCCAGGATTCAAAATTCATGTGTAGGAT 1054
 |||||:||||| ||| :||| |||||
 Db 366 rSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro-ValThr 384
 |||||:||||| ||| :||| |||||
 QY 1055 TTAGACCAAGAAAGAAAGATGGCTCAAGCTCCACAGAGATGCACTTACGGCATCT 1114
 |||||:||||| ||| :||| |||||
 Db 385 -----SerSerThrThrGluSerSerSerAlaProValP 396
 |||||:||||| ||| :||| |||||
 QY 1115 TTAGACACACAGTGCAGAGCAAAAGCCCTGCAAGTGCACCTCTCTTGTATCCA 1174
 |||||:||||| ||| :||| |||||
 Db 396 roThrProSerSerSerThrThrGluSerSerSerAlaProValThrSer---SerThr 415
 |||||:||||| ||| :||| |||||
 QY 1175 ACAAAATGAAAGTCCAGGAGTCTATCATGACCATCGAGAGGAGAGCAACAGAGAA 1234
 |||||:||||| ||| :||| |||||
 Db 415 hrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 435
 |||||:||||| ||| :||| |||||
 QY 1235 TCTATCTCAGCTACAGACCTCAAAAGGCTGATCAGCAAGCACTAGAGAGAGAACAT 1294
 |||||:||||| ||| :||| |||||
 Db 435 hrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 455
 |||||:||||| ||| :||| |||||
 QY 1295 CTTTGGATGGGGCAATTCAGTTCATGATGAATTCGTGGATCACTGCCACCTTTG 1354
 |||||:||||| ||| :||| |||||
 Db 455 erAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrS 475
 |||||:||||| ||| :||| |||||

QY 1355 GTCTGACACCCAAATCAGAG-----CTGCCCAACA-----1383

Db 475 erSerThrThrGlusSerSerAlaProValProThrProSerSerThrThrGlus 495

QY 1384 --TCCTTTTGCTTATACAGAGGATGCCTACTTTGAGTCCAGAACCCTCTCTCGTTGGAAC 1441

Db 495 erSerSerAlaProValThrSerSerThrThrThrGlusSerSerAlaProValProThr 515

QY 1442 CCCAGCTTGAGACAGTGGACGGAGCAGAGCATGGCTCACT-----G 1483

Db 515 roSerSerThrThrThrGlusSerSerAlaProAlaProThrProSerSerSerThrTh 535

QY 1484 ACACCTTTGTCTCCACCTGCTATGGCTCTACTCCCTGTCAGAGCTCCACACTTCT 1543

Db 535 hrGlusSerSerAlaProValThrSerSerThrThrThrGlusSerSerAlaProValP 555

QY 1544 TTATGGCATCAAGCATCTTCTCTCACTGATCMAGGCACACAGAT--ACAATTGGCCA 1600

Db 555 roThrProSerSer-----SerThrThrThrGlusSerSerThrProValThrSerSert 573

QY 1601 CTGACGACGAATCTAGTACCAGGCGCTCACCACTCCCCACCAAGTGATTAATCTGCACATCA 1660

Db 573 hrThrGlusSerSerSerAlaPro-----ValProThrProSerSerSerThrThrG 590

QY 1661 GCCAAGCTGCTCGGAAATTCATATCCACTGCACTCTTCAGATGACACGCCGATCAAGTG 1720

Db 590 luSerSerSerAlaProValProThrProSerSerSerThrThrThrGlus-----606

QY 1721 CAGGTGGCGAAGATATGGTCAGACACCTAGATAAATGATGCTGCTGACACTCTGCCC 1780

Db 607 -----SerSerAlaProAlap 612

QY 1781 CATCTGAGTACACAGCTCAGCGAATATGTTCTGCCAGATCATTTCTTGGAGGATA 1840

Db 612 roThrProSerSerSerThrThrThrGlusSerSerSerAlaPro-----ValThrSerSer 630

QY 1841 CCACCT-----CCTCTCTCAGCTTACAGATATATCACT-----AGTT 1879

Db 630 hrThrGlusSerSerSerAlaProValProThrProSerSerSerThrThrThrGlusSer 650

QY 1880 CTATGACCATGCCC---CCAAAGGCGAGAGCTGGTAGTGTCTTCACTGCTGGGTTG 1936

Db 650 erAla-ProValProThrProSerSerSerThrThrThrGlusSerSerAlaProVal---668

QY 1937 CTAAATGCGCTTCCACACAGCTGTTCCACAGAGCTCTTGGAGTACCGAGCTCTGG 1996

Db 669 -----ProThrProSerSerSerThrThrThrGlusSerSerSerAlaProValThrSe 685

QY 1997 AGCAACAAATTCACAGCTGCTGTTCCATA-----TCTACGATCCAACT 2042

Db 685 rSerThrThrThrGlusSerSerSerAlaProValThrSerSerSerThrThrThrGlus 702

RESULT 15

S61535
nucleotide-binding head-stalk protein 183K - Giardia lamblia
C:Species: Giardia lamblia
C>Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
F:Marshall, J.; Holberton, D.V.
J. Cell Sci. 108, 2683-2692, 1995
A:Title: Giardia gene predicts a 183 kDa nucleotide-binding head-stalk protein.
A:Reference number: S61535; MUID:96066038; PMID:7593309
A:Accession: S61535
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1620 <M>

| | | |
|------------------------|--------|---------------|
| Alignment Scores: | | |
| Pred. No.: | 0.106 | length: |
| Score: | 147.00 | Matches: |
| Percent Similarity: | 34.50% | Conservative: |
| Best Local Similarity: | 18.53% | Mismatches: |
| | | |

| | | | |
|--------------|-------|---------|-----|
| Query Match: | 2.50% | Indels: | 270 |
| DB: | 2 | Gaps: | 35 |

US-10-007-270-1 (1-3330) x S61535 (1-1620)

```

QY   222 CTGAACTTAAGACATAGACAATCCCCCAAGAATGAACAACACTGAAGTACTGAAAAA 281
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 fllysgluylserValThrIleGluLysLeuMetGluLysLeuSerLySleuLys 332
QY   282 TGTCAAATGTCAACTATGAGCAGNATTTCGATTGGCAAGCATCGAACACAAAAGAT 341
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
333 ThrSerGlnVal-----AlalleuglInGasp 342
QY   342 CGGCATTTTCCCAA-----CGGGGGTTAAAGTCTGTCCACAGG----- 380
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
343 GlnAspAlaValGluLeuAlaArgSerValallaaspMetArgGlnLySleuSerAspThr 362
QY   381 ---AATCCATGAACAGNTTTAGACAGTCTTCAAGCTTTATTATAGATTGAGAGTGTC 437
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
363 ValaspluLysGlnArglle-----ValHisGlnLeuHisileaspMetAspSerMet 380
QY   438 AGGAACGAGTATGGAAGCATATCGGATCTTTC----- 470
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 LysLySerHisGluSerIleSerGlnHisLySerGluIleGluAlalleLysSer 400
QY   470 ----- 470
Db    ----- 470
401 AlaPheAspGluSerThrPheIleLysAsnAlaArgIleSerLySleuThrThrGluLeu 420
QY   471 -----TGGATCCGATCCCTGCACACAGGGGATATCACAGCTGGGTGACGA 515
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 ThrGluThrGlnThrGlnLeuAlaSerAlaLeuGluLysAsnGlnGluLeuGlnSerGlu 440
QY   516 TCTGCCACGACGAGACCTTCTGCTCTTTCACATTTGAAAAAATCTCAGCAATCCCAGG 575
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 SerLySleuAlaGlnHisalalaagluValLeuLeuAlaGluArgGlu---llealaasn 459
QY   576 AGCACTCGGATCTTCTCCAGCAGAGATAAAACAGAGAAGTTTCCCTGACAGAAAGATG 635
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
460 ThrThrLeuThrGlnLeuGlnAlaThrLySleGlyGluLeuGluAlaValArgLySer 479
QY   636 AATATCTGCAGAGAAGACATTCGACAGCTGTGTAACCCTCAGAAAGCAATTCAGA 755
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
480 AsnHisGluGlnValAsnSerTyGluSerLeuLeuGln-----LysThrGln 495
QY   696 TCTCATTTCAAGACTTGGCGCATATCTTAAGAAACCTCAGAAAGACAAATTCAGA 755
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
496 GlnSerTyrglualeuLeuArgGluLysGluAspThrIleLeuLysleuasnLeuGlu 515
QY   756 TGTTGCCAAGCTCTCACTTGGGCTTTCCCTCTCACTCTCTGATGACACCCCTCTCAATGA 815
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 ---CyaspAsnalalyeAlaPheAspAspHisSer----- 527
QY   816 AATTCGTGATATACATCAACGACACCAAGATCCCTACACAGAAAGACAGAAACAGAATT 875
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
528 -----GluGlnSerThrAlaHisValGlnAlaLeuArgGluValGluArgVal 544
QY   876 CGCTGTGTGGAGGACGAGAGGTGGAGCTCAGCGTCTCTCTGGTAAAC----- 924
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 ArgHisThr-SerGluasnLeuGlnGluLysMetGlyValilleasnargLeuThrAl 564
QY   925 -----CAGAAGTCCAAG----- 936
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
564 alGluIleGlnAlalleLysleuAspGlyLuarGlaLeuAspSerLySleuAspThrGlnIle 584
QY   937 -GAGAGACTCGCTCACTCCGACTCCCATATATTACAGGAGCTA-----GCAGAAA 986
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
584 eAlaGluLeuGlnLysasnValSerSerLeuMetGluaspLeuThrLySleuSerSerAs 604
QY   987 GTCCCAACTTCAGATGCAAAAGATATTTAAGAACTTCCAGGATTCAAAANAATCCATGT 1046
Db    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 nSerSerAlaGluLeuSerArgLeuThrAsnLeuLeuGluAlaThrArgLyAspTyVa 624

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 4, 2004, 18:37:36 ; Search time 32.5 Seconds
(without alignments)

10670.376 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taacacaagaaggttatcct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2_1/USPTO_pool_p/US10007270/runat_04032004_160716_4912/app_query.fasta_1.3527
-DB=SwissProt_42 -OFT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10007270 @cgn 1 1 36 @runat_04032004_160716_4912 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--|
| 1 | 189.5 | 3.2 | 1255 | 1 MUC1 HUMAN | P15941 h mucin 1 p |
| 2 | 187.5 | 3.2 | 475 | 1 MUC1 HYLA | Q29435 hylobates 1 |
| 3 | 153 | 2.6 | 3678 | 1 DMD_MOUSE | P11531 mus musculus |
| 4 | 152.5 | 2.6 | 3358 | 1 PGCV_MOUSE | Q62059 mus musculus |
| 5 | 152.5 | 2.6 | 8797 | 1 SNEI_HUMAN | Q8nf91 homo sapien |
| 6 | 151 | 2.6 | 1341 | 1 ACIN_HUMAN | Q9ukj3 homo sapien |
| 7 | 150 | 2.6 | 1163 | 1 RTN4_RAT | Q9jkl8 rattus norv |
| 8 | 149 | 2.5 | 7389 | 1 BPAL_MOUSE | Q9izt6 mus musculus |
| 9 | 148 | 2.5 | 1490 | 1 CRK7_HUMAN | Q9nyv4 homo sapien |
| 10 | 148 | 2.5 | 2738 | 1 PGCV_RAT | Q9erB4 rattus norv |
| 11 | 147 | 2.5 | 1367 | 1 AMYH_YEAST | P08640 saccharomyc |
| 12 | 146.5 | 2.5 | 3396 | 1 PGCY_HUMAN | P13611 homo sapien |
| 13 | 145.5 | 2.5 | 928 | 1 NIBA_HUMAN | RC MEDLINE=90368715; PubMed=1697589; |
| 14 | 145.5 | 2.5 | 1024 | 1 POPC_FALSO | Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T., |
| 15 | 145 | 2.5 | 630 | 1 MUC1_MOUSE | Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D., |
| 16 | 145 | 2.5 | 1036 | 1 P200_MYCPN | "Molecular cloning and expression of human tumor-associated |
| 17 | 145 | 2.5 | 1957 | 1 SPOF_SCHPO | polymorphic epithelial mucin." |
| 18 | 145 | 2.5 | 2869 | 1 RBPI_PLAVB | J. Biol. Chem. 265:15286-15293(1990). |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 19 | 144 | 2.5 | 1468 | 1 N153_RAT | P49791 rattus norv |
| 20 | 143.5 | 2.4 | 1062 | 1 YAI3_HUMAN | Q9Y216 homo sapien |
| 21 | 143.5 | 2.4 | 1306 | 1 MSB2_YEAST | P23234 saccharomyc |
| 22 | 143.5 | 2.4 | 3680 | 1 DMD_CANFA | Q97592 canis faml |
| 23 | 143.5 | 2.4 | 3685 | 1 DMD_HUMAN | P11532 homo sapien |
| 24 | 143 | 2.4 | 3259 | 1 GOB1_HUMAN | Q14789 homo sapien |
| 25 | 141.5 | 2.4 | 3381 | 1 PGCY_BOVIN | P81282 bos taurus |
| 26 | 141 | 2.4 | 2845 | 1 APC_MOUSE | Q61315 mus musculu |
| 27 | 140.5 | 2.4 | 529 | 1 HSF1_HUMAN | Q00613 homo sapien |
| 28 | 140.5 | 2.4 | 1772 | 1 MSP1_PLAYO | P13828 plasmodium |
| 29 | 139.5 | 2.4 | 1928 | 1 MYSL_YEAST | P08964 saccharomyc |
| 30 | 139 | 2.4 | 2145 | 1 CYAA_PODAN | Q01513 podopoxa a |
| 31 | 139 | 2.4 | 5120 | 1 PCLO_CHICK | Q9np36 gallus gall |
| 32 | 139 | 2.4 | 8545 | 1 ANCI_CAEEL | Q9n4m4 caenorhabdi |
| 33 | 138.5 | 2.4 | 2230 | 1 GOA4_HUMAN | Q13439 homo sapien |
| 34 | 138.5 | 2.4 | 4128 | 1 PRKD_MOUSE | P97313 mus musculu |
| 35 | 138 | 2.3 | 1677 | 1 VIT_ACITR | Q90243 acipenser t |
| 36 | 137.5 | 2.3 | 1609 | 1 FIG2_YEAST | P25653 saccharomyc |
| 37 | 137.5 | 2.3 | 6885 | 1 SNE2_HUMAN | Q8wxh0 homo sapien |
| 38 | 137 | 2.3 | 1876 | 1 APSA_EMENI | Q00083 emericeila |
| 39 | 136.5 | 2.3 | 2158 | 1 MY9B_HUMAN | Q13459 homo sapien |
| 40 | 135.5 | 2.3 | 917 | 1 SMOO_HUMAN | P53814 homo sapien |
| 41 | 135.5 | 2.3 | 1142 | 1 GIN4_YEAST | Q12263 saccharomyc |
| 42 | 135.5 | 2.3 | 1222 | 1 YMH5_CAEEL | P34472 caenorhabdi |
| 43 | 135 | 2.3 | 1140 | 1 YMG6_YEAST | Q04893 saccharomyc |
| 44 | 135 | 2.3 | 1464 | 1 NCO2_HUMAN | Q15596 homo sapien |
| 45 | 135 | 2.3 | 1875 | 1 MLPI_YEAST | Q02455 saccharomyc |

ALIGNMENTS

| | | | |
|----------|---|-----------|---------------|
| RESULT 1 | MUC1_HUMAN | STANDARD; | PRT; 1255 AA. |
| ID | MUC1_HUMAN | STANDARD; | PRT; 1255 AA. |
| AC | P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442; | | |
| AC | Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQU1; Q9Y4J2; | | |
| DT | 01-JAN-1990 (Rel. 13, Created) | | |
| DT | 01-APR-1990 (Rel. 14, Last sequence update) | | |
| DT | 10-OCT-2003 (Rel. 44, Last annotation update) | | |
| DE | Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT) | | |
| DE | (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin) | | |
| DE | (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut- | | |
| DE | reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen | | |
| DE | DF3) (CD227 antigen). | | |
| GN | MUC1. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| OX | NCBI_TaxID=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. (ISOFORM 1). | | |
| RC | TISSUE=Pancreas; | | |
| RX | MEDLINE=90368716; PubMed=2394722; | | |
| RA | Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.; | | |
| RT | *Cloning and sequencing of a human pancreatic tumor mucin cDNA.*; | | |
| RL | J. Biol. Chem. 265:15294-15299(1990). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. (ISOFORMS 1 AND 2). | | |
| RC | MEDLINE=90202794; PubMed=2318825; | | |
| RA | Ligtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.; | | |
| RT | *Episialin, a carcinoma-associated mucin, is generated by a | | |
| RT | polymorphic gene encoding splice variants with alternative amino | | |
| RT | termini.*; | | |
| RL | J. Biol. Chem. 265:5573-5578(1990). | | |
| RN | [3] | | |
| RP | SEQUENCE FROM N.A. (ISOFORM 1). | | |
| RC | TISSUE=Breast carcinoma; | | |
| RX | MEDLINE=90368715; PubMed=1697589; | | |
| RA | Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T., | | |
| RT | Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.; | | |
| RT | "Molecular cloning and expression of human tumor-associated | | |
| RT | polymorphic epithelial mucin." | | |
| RL | J. Biol. Chem. 265:15286-15293(1990). | | |

RN [4] SEQUENCE FROM N.A. (ISOFORM 1).
 RP MEDLINE=91097524; PubMed=2268309;
 RX Lancaster C.A., Peat N., Duhig T., Wilson D.,
 RA Taylor-Papadimitriou J., Gendler S.J.;
 RT "Structure and expression of the human polymorphic epithelial mucin
 gene: an expressed VNTR unit.";
 RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 5).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=90276413; PubMed=2351132;
 RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
 RA Zaretzky J., Kotkes P., Weiss M., Lathe R., Keydar I.,
 RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
 may generate multiple protein forms.";
 RL Eur. J. Biochem. 189:463-473(1990).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=90276414; PubMed=2112460;
 RA Hareuveni M., Tsarfaty I., Zaretzky J., Kotkes P., Horev J.,
 RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
 RT "A transcribed gene, containing a variable number of tandem repeats,
 codes for a human epithelial tumor antigen. cDNA cloning, expression
 of the transfected gene and over-expression in breast cancer
 tissue.";
 RL Eur. J. Biochem. 189:475-486(1990).
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=91033045; PubMed=1688329;
 RA Tsarfaty I., Hareuveni M., Horev J., Zaretzky J., Weiss M.,
 RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
 RT "Isolation and characterization of an expressed hypervariable gene
 coding for a breast-cancer-associated antigen.";
 RL Gene 93:313-318(1990).
 RN [8]
 RP SEQUENCE FROM N.A. (ISOFORM 7).
 RX MEDLINE=95010060; PubMed=7925397;
 RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
 RA Keydar I., Hilkens J., Wreschner D.H.;
 RT "Characterization and molecular cloning of a novel MUC1 protein.
 devoid of tandem repeats, expressed in human breast cancer tissue.";
 RL Eur. J. Biochem. 224:787-795(1994).
 RN [9]
 RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
 RX MEDLINE=97355747; PubMed=9212228;
 RA Osterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
 RA Finstad C.B.;
 RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
 cancer cell lines and demonstration of a new short variant form
 (MUC-1/2).";
 RL Int. J. Cancer 72:87-94(1997).
 RN [10]
 RP SEQUENCE FROM N.A. (ISOFORM 7).
 RA Zhang L.X., Li C.H.;
 RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [11]
 RP SEQUENCE FROM N.A. (ISOFORM 9).
 RC TISSUE=Epithelial cancer;
 RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;
 RT "Cloning of a new potential secreted short variant form of MUC1 mucin
 in epithelial cancer cell line.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=88330762; PubMed=3417635;
 RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
 RA Burchell J.;
 RT "A highly immunogenic region of a human polymorphic epithelial mucin
 expressed by carcinomas is made up of tandem repeats.";
 RL J. Biol. Chem. 263:12820-12823(1988).
 RN [13]
 RP SEQUENCE OF 1-150 FROM N.A. (ISOFORM 2).
 RX MEDLINE=9008473; PubMed=2597151;
 RA Abe M., Siddiqui J., Kufe D.;
 RT "Sequence analysis of the 5' region of the human DF3 breast
 carcinoma-associated antigen gene.";
 RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
 RN [14]
 RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
 RC TISSUE=Thyroid;
 RX MEDLINE=96183746; PubMed=8608966;
 RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
 RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
 transcriptase polymerase chain reaction of the MUC1 gene.";
 RL Int. J. Cancer 66:55-59(1996).
 RN [15]
 RP SEQUENCE OF 1-89 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=96181716; PubMed=8604237;
 RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
 RA Lee L.N., Luh K.T., Wu C.W.;
 RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
 tissues.";
 RL Oncology 53:118-126(1996).
 RN [16]
 RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Breast carcinoma;
 RA Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=97460054; PubMed=9312074;
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
 RA Hanisch F.-G.;
 RT "Localization of O-glycosylation sites on glycopeptide fragments from
 lactation-associated MUC1. All putative sites within the tandem
 repeat are glycosylation targets in vivo.";
 RL J. Biol. Chem. 272:24780-24793(1997).
 RN [18]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=99303572; PubMed=10373415;
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
 RA Hanisch F.-G.;
 RT "High density O-glycosylation on tandem repeat peptide from secretory
 MUC1 of T47D breast cancer cells.";
 RL J. Biol. Chem. 274:18165-18172(1999).
 RN [19]
 RP POLYMORPHISM WITHIN THE REPEAT.
 RX MEDLINE=97459366; PubMed=11350974;
 RA Engelmann K., Balduz S.E., Hanisch F.-G.;
 RT "Identification and topology of variant sequences within individual
 repeat domains of the human epithelial tumor mucin MUC1.";
 RL J. Biol. Chem. 276:27764-27769(2001).
 RN [20]
 RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
 RX MEDLINE=99211485; PubMed=10197628;
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereth Y., Greenstein S.,
 RA Stadler Y., Skornik Y., Zaretzky J., Smorodinsky N.I., Keydar I.,
 RA Wreschner D.H.;
 RT "The breast cancer-associated MUC1 gene generates both a receptor and
 its cognate binding protein.";
 RL Cancer Res. 59:1552-1561(1999).
 RN [21]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
 RX MEDLINE=21240104; PubMed=11341784;
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
 RA Harris A.;
 RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
 RN [22]
 RP CHARACTERIZATION.
 RX MEDLINE=21836452; PubMed=11847293;
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,

| | | | |
|------|----|--|------|
| 1345 | QY | CCAGCGCTTGGT-----CCTGACCCCAATCAGAGCTGCCACATCTTTTGGT | 1392 |
| 814 | Db | ProAlaHisGlyValThrSerAlaProAspThrArg---ProAlaProGlySerThrAla | 832 |
| 1393 | QY | GTATATACAGAGGATGCTACTTTGAGTCCAGACATCTCTCCTGTGTGAACCCAGCTTGAG | 1452 |
| 833 | Db | ProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly-----Ser | 850 |
| 1453 | QY | ACAGTGCAGCAGCAGACATGCTCTA-----CCTGCACACT----- | 1488 |
| 851 | Db | ThrAlaProProAla---HisGlyValThrSerAlaProAspThrArgProAlaProGly | 869 |
| 1489 | QY | TCCTGGTCTCCACTGCTGATGGCCTCTACCTCCCTG----- | 1524 |
| 870 | Db | SerThrAlaProProAlaHisGlyValThrSerAlaProAspThrArgProAlaProGly | 889 |
| 1525 | QY | TCAGAGCTCCACCTTTCTTATGGCATCAAGCATCTTCTCTGACTGAT----- | 1575 |
| 890 | Db | SerThrAlaProPro-----AlaHisGlyValThrSerAlaProAspThrArgPro | 906 |
| 1576 | QY | ---CAAGGCACACAGAT-----ACATGGCCACTGACACAGCAATG | 1614 |
| 907 | Db | AlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAsp----- | 925 |
| 1615 | QY | CTAGTACAGGGCTCCACATCCCCACAGT-----GATTATCTGCAATCAGCCAA | 1665 |
| 926 | Db | ProAlaProGlySerThrAlaProProAlaHisGlyValThrSerAlaProAspAsnArg | 945 |
| 1666 | QY | CTGGCTCTGGGAATTCACATCCACT-----GCATCTTCAGATGACAGCGCATCA | 1716 |
| 946 | Db | ProAlaLeuGlySerThrAlaProProValHisAsnValThrSerAlaSerGlySerAla | 965 |
| 1717 | QY | AGTGCAGGTGGCAAGATATGCTCAGACACTAGATGAATGATCTGTCGACACTCT | 1776 |
| 966 | Db | SerGlySerAlaSerThrLeuValHisAsnGlyThrSerAlaArgAlaThrThrPro | 985 |
| 1777 | QY | GCCCATCTGAGGTACAGAGCTCAGCGAATATGTTCTGCCAGATCATCTTCTGGAG | 1836 |
| 986 | Db | AlaSerLysSerThrPro-----PheSerIleProSerHisHisSerAsp | 1000 |
| 1837 | QY | GATACCACTCTCTCTC----- | 1854 |
| 1001 | Db | ThrProThrThrLeuAlaSerHisSerThrLysThrAspAlaSerSerThrHisHisSer | 1020 |
| 1855 | QY | GCTTTACAGTATATACCACTAGTCTCTATGACATTGCCCCAGGGCCGAGAGCTGGTA | 1914 |
| 1021 | b | SerValProProLeuThrSerSerAsnHisSerThrSerProGlnLeuSerThrGlyVal | 1040 |

```

FT REPEAT 122 141 2.
FT REPEAT 142 161 3.
FT REPEAT 162 181 4.
FT DOMAIN 254 371 SEA.
SQ SEQUENCE 475 AA; D7A699D6D68C6622 CRC64;

Alignment Scores:
Pred. No.: 0.000324 Length: 475
Score: 187.50 Matches: 85
Percent Similarity: 39.60% Conservative: 52
Best Local Similarity: 24.57% Mismatches: 124
Query Match: 3.19% Indels: 85
DB: 1 Gaps: 17

US-10-007-270-1 (1-3330) x MUC1_HYLLA (1-475)

QY 1291 CAATCTTGGATGCGGACCAATTCAGTTCATGATGAATTCGTGATCATCGCCAGCC 1350
Db 74 GlnGlyGlnAspValSerLeuAlaProAlaThrGluProAlaSerGlySerAlaAlaThr 93
QY 1351 TTTCGTCTCCAGACCAATCA-----GAGCTGCCACATCTTTGCTGTT 1395
Db 94 TrpGlyGlnAspValThrSerValProValThrArgProAlaProGlySerThrThrSer 113
QY 1396 ATACACAGAGCTACTTTCAGTCCAGAACTTCCTCTGTGAACCCAGCTTGACACA 1455
Db 114 ProAlaGlnAspValThrSerAlaProAsp-----ThrArgProAlaLeuGlySer 130
QY 1456 GTGACGGGAGCAGACAGTGTCTA-----CTGACACTTCTTGGTCTCCACTGCT 1506
Db 131 ThrAlaProProValHisGlyValThrSerAlaProAspThrArg-----ProThr 147
QY 1507 ATGCGCTTACCTCCCTCTCAGAGCTCCACCTTCTTATGCGATCAGCATCTCTCT 1566
Db 148 LeuGlySerThr-----AlaProProVal----- 155
QY 1567 CTGACTGATCAAGCCACACAGATACATGCGCACTGACACAGACAATCTAGTACCAGG 1626
Db 156 -----HisGlyValThrSer-----AlaProAsp----- 163
QY 1627 CTCACATCCACACAGATGATTTCTGCAATCAGCAACTGGCTCTGGATTTCCAT 1686
Db 164 ---ThrArgProThr-----LeuGlySerThrAla 172
QY 1687 CCACCT-----GCATCTTCAGATGACAGCGATCAAGTGCAGTGGCGAGATAG 1737
Db 173 ProProValHisAsnValThrSerAlaSerGlySerGlySerAlaSerThrLeu 192
QY 1738 GTCAGACCTAGATGAATGGATCTGTCTGACACTCTGCCCCCACTGAGGTACCAGAG 1797
Db 193 ValHisAsnGlyThrSerAlaArgAlaThrThrProAlaSerLysSerThrPro--- 211
QY 1798 CTCAGCGAATATGTTCTGCCAGCATATTTCTGGAGGATACCATCTCTGTC----- 1851
Db 212 -----PheSerIleProSerHisHisSerAspThrProThrThrLeuThrSer 227
QY 1852 -----TCAGCTTTACAGTATATACCACT 1875
Db 228 HisSerThrLysThrAspAlaSerSerThrHisHisSerThrValSerProLeuThrSer 247
QY 1876 AGTTCTATGACATTCGCCCAAGGCGGAGAGCTGTAGTGTCTTCTC-----AGTCTG 1929
Db 248 SerAsnHisSerThrSerProGlnLeuSerIleGlyValSerPhePheLeuSerPhe 267
QY 1930 CGTGTGTACATGCGCTTCTCCAGCCTGTTCCACAGCTTTCACAGAGCTCTCGAGTACGA 1989
Db 268 HisIleSerAsnLeuGlnPheAsnSerSerLeuGluAspProSerThrAsnTrpTrpGln 287
QY 1990 GCTCTGGAGCAACAAATTCACAGCTGTGTTCCA-----TATCTACCATTCATCA 2046
Db 288 GluLeuGlnArgAspIleSerGluLeuIleGluGlnIleTrpLysGlnGlyAspPheLeu 307
QY 2047 GGATTAAAGCACTTGAATACTTAATCTTCAGAAACGGAGTGTGATTGTAATAGCAA 2106

```

```

Db 308 GlyValSerAsnIleLys-----PheArgProGlySerValValGlnSerThr 324
QY 2107 ATGAAGTTTGTAGTCTGTGCGGTATTAACCTCCACAGGCTGTGCGACGGGCTTTGGAG 2166
Db 325 LeuAlaPheArgGluGly-----ThrAsnValHisAspValGluAla 339
QY 2167 GATTTT-----CGTTCTGCTGACGCCCAACACTCCATCTCGAAATAGACAGCTAC 2217
Db 340 GlnPheAsnGlnHisLysThrGluAlaSerArgTyrAsnLeuThrIleSerAspVal 359
QY 2218 TCTTCAACATTTGAACCA 2235
Db 360 SerValSerAspValPro 365

RESULT 3
DMD_MOUSE
ID DMD_MOUSE STANDARD; PRT; 3678 AA.
AC P11531; O35653; Q60703;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Dystrophin.
GN DMD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=92253376; PubMed=1579466;
RA Bies R.D.; Phelps S.F.; Cortez M.D.; Roberts R.; Caskey C.T.;
RA Chamberlain J.S.;
RT "Human and murine dystrophin mRNA transcripts are differentially
RT expressed during skeletal muscle, heart, and brain development.";
RL Nucleic Acids Res. 20:1725-1731(1992).
RN [2]
RP SEQUENCE OF 1-201 FROM N.A.
RP MEDLINE=87273512; PubMed=3607877;
RA Koenig M.; Hoffman E.P.; Bertelson C.J.; Monaco A.P.; Feener C.;
RA Kunkel L.M.;
RT "Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
RT preliminary genomic organization of the DMD gene in normal and
RT affected individuals.";
RL Cell 50:509-517(1987).
RN [3]
RP SEQUENCE OF 120-176 FROM N.A.
RP STRAIN=129/J;
RP MEDLINE=92182520; PubMed=1543903;
RA Maconochie M.K.; Brown S.D.M.; Greenfield A.J.;
RT "Sequence analysis of two exons from the murine dystrophin locus.";
RL Mamm. Genome 2:64-68(1992).
RN [4]
RP SEQUENCE OF 300-1390 FROM N.A.
RP MEDLINE=88018015; PubMed=3659917;
RA Hoffman E.P.; Monaco A.P.; Feener C.C.; Kunkel L.M.;
RT "Conservation of the Duchenne muscular dystrophy gene in mice and
RT humans.";
RL Science 238:347-350(1987).
RN [5]
RP SEQUENCE OF 986-1056 FROM N.A.
RP STRAIN=C57BL/10; TISSUE=Skeletal muscle;
RX MEDLINE=94154933; PubMed=8111539;
RA Chamberlain J.S.; Phelps S.F.; Cox G.A.; Maichele A.J.;
RA Greenwood A.D.;
RT "PCR analysis of muscular dystrophy in mdx mice.";
RL Mol. Cell Biol. Hum. Dis. Ser. 3:167-189(1993).
RN [6]
RP ALTERNATIVE SPLICING.
RP STRAIN=C57BL/10; TISSUE=Retina;
RX MEDLINE=95360002; PubMed=7633443;
RA D'Souza V.N.; Nguyen T.M.; Morris G.E.; Karges W.; Pillers D.-A.M.;
RA Ray P.N.;

```


Db 2088 GluLysLeuHisArgMetTyrLysGluArgGlnGlyArgPheAspArgSerValGluLys 2107
 QY 485 -----GACACAGGGGAATATCAGGACTGGGTGAGCTCTGAGCAGAG 526
 Db 2108 TrpArgHisPheHisTyrAspMetLysValPheAsnGlnTrpLeuAsnGluValGluGln 2127
 QY 527 GAGACCTTCTGCTCTTGGATGGTGAATAAATCTCAGCAATTCACAGAG- 577
 Db 2128 -----PhePheLysLysThrGlnAsnProGluAsnTrpGluHisAlaLysTyr 2143
 QY 578 -----CACCTGATCTTCTCCAGACAGATAAACAAGAGAGTTTCCCTGACACAAA 631
 Db 2144 LysTrpTyrLysGluLeuGlnAspGlyIleGlyGlnArgGln- 2158
 QY 632 GATGAATATCTGACAGAGACATTTGGAGCGCTGGTGAACACCATTTGTCATTTCAACA 691
 Db 2159 -----AlaValValArgThrLeuAsnAlaThrGlyGluGluIleGlnGlnSer 2175
 QY 692 GC-----AATCTACATTTCAAGACTTGGCGAGTATT- 723
 Db 2176 SerLysThrAspValAsnIleLeuGlnLysLeuGlySerLeuArgTrpHis 2195
 QY 724 -----CTAAGAAAACCTCAGAAAGACAA- 747
 Db 2196 AsnIleCysLysGluLeuAlaGluArgArgLysArgIleGluGluGlnLysAsnValLeu 2215
 QY 748 -----ATCAAGATGTTGCCAACGTC 768
 Db 2216 SerGluPheGlnArgAspLeuAsnGluPheValLeuTrpLeuGluGluAlaAspAsnIle 2235
 QY 769 TCACCTGGGCTTTCCTCTCTCCTGATGACACCTCCTCAATGAAATCTCGAAT 828
 Db 2236 AlaIleThrProLeu-----GlyAspGluGlnLysLeuLysGlu- 2248
 QY 829 ACATCTCAACGACACCAAGATGCTACACAGAAAGAGAAACAGAAATTCGCTGTGTGAG 888
 Db 2249 GlnLeuGluGlnValLysLeuAlaGluGluLeuPLeuArgGlnGlyIleLeuLys 2268
 QY 889 GAGCAGAGGTGGAGCTCAGCGTCTCTCTGTAACACGAAATTCACGACGAGTGTGCT 948
 Db 2269 GlnLeuAsnGluThrGlyGlyAlaValLeuValSerAlaProIleArgProGluGluGln 2288
 QY 949 GACTCCCGAGTCCCAATATTACAGGAGCTAGCAGGAAAGTCCCACTTCAGATGCAAAAG 1008
 Db 2289 AspLys-----LeuGluLysLysLeuLysGlnThrAsnLeuGlnTrpIleLys 2304
 QY 1009 ATATTAAAGAACTTCCA-----GGATTCAAAAAATCCATGTGTAGGATTTAGA 1059
 Db 2305 ValSerArgAlaLeuProGluLysGlnGlyGluLeuGluValHisLeuLysAspPheArg 2324
 QY 1060 CCAAGAAAGAAAGATGGCTCAAGCTCCAGAGATGCAACTTACGGCCATCTTTAAG 1119
 Db 2325 GlnLeuGluGln-----LeuAspHisLeuLeuLeuTrpValSerProIle 2340
 QY 1120 AGACACAGTGCAGAGCAAAAGCCCTGCAGTACCTCTGCTCTTTGATTCACACAA 1179
 Db 2341 ArgAsnGlnLeuGluIleTyrAsnGlnProSerGlnAlaGlyPropPheAspIleLysGlu 2360
 QY 1180 ATTCAAGTGCAGAGTCTATCATGGAACCATGGAGGACAGCAACACCAAAATCTAT 1239
 Db 2361 Ile-----GluValThrValHisGly-----LysGln- 2369
 QY 1240 CTCACAGCTACAGACTCAAAAGCTGATCAGCAAGCA-----CTAGAGGAGAACAA 1293
 Db 2370 -----AlaAspValGluArgLeuLeuSerLysGlyGlnHisLeuTyrLysGluLys 2386
 QY 1294 TCTTTGGATGTGGGACAAATTCAGTCTCATGATCAATTCGTTGGATCACTCCAGCCTTT 1353
 Db 2386 ----- 2386
 QY 1354 GGTCTCTGACCCCAATCAGAGCTGGCCCACTCTTTTCTGTTTATTAACAGAGGATGCTACT 1413
 Db ----- 2386

Db 2387 ---ProSerThrGln----- 2390
 QY 1414 TTGAGTCCAGAACTTCTCTCTGTTGAACCCAGCTTGAGCAGTG-----GAC 1461
 Db 2391 -----ProValLysArgLysLeuGluAspLeuArgSerGlnTrpGlu 2404
 QY 1462 GGAGCAGAGATGGTCTA-----CTCAGCACTTCTTGCTCTCCA 1500
 Db 2405 AlaValAsnHisLeuLeuArgGluLeuArgThrLysGlnProAspArgAla- 2421
 QY 1501 CCTCTATGGCTCTTACCTCCCTGTCAGAGCTCCACCTTCTTATATGCTTCAAGATC 1560
 Db 2422 ProGlyLeuSerThrThrGlyAlaSer-----AlaSerGlnThr 2434
 QY 1561 TTCTCTCTGACTGATCAGGACCAAGGACCAAGATACAAATGAGCCACTGACGACAACTCTAGTA 1620
 Db 2435 ValThrLeu-----ValThrGlnSerValValThrLysGluThr---ValIle 2449
 QY 1621 CCAGGGCTCACCATCCCACTGATGATTATTCGATCAGCACTGGCTCTGGGAAT 1680
 Db 2450 SerLysLeuGluMetProSerSer-----LeuLeuLeuGluVal 2462
 QY 1681 TCACATCCACTGATCTTTCAGATGACGCGATCAAGTCAGAGTGGCGAGATATGTC 1740
 Db 2463 -----ProAlaLeuAlaAspPheAsnArgAlaTrpThrGluLeuThrAspTrpLeu 2479
 QY 1741 AGACACTAGATGAATGGATCTGTCTGACACTCTGCCCCATCTGAGTACACAGAGCTC 1800
 Db 2480 SerLeuLeuAspArgValIleLysSerGlnArgValMetValGlyAspLeuGluAspIle 2499
 QY 1801 AGCGAATATGTT-----TCTGTCCAGATCATCTTCTGGAGGATACCACT 1845
 Db 2500 AsnGluMetIleIleLysGlnLysAlaThrLeuGlnAsp-----LeuGluGlnArg 2517
 QY 1846 CTGTCTCAGCTTTACAGTATATCACCAGTATCTATCAGCATTCGCCCCAG----- 1899
 Db 2518 Pro-----GlnLeuGluGluLeuIleThrAlaAlaGlnAsnLeuLysAsnLysThrSer 2535
 QY 1900 GGCGAGAGCTGGTAGTGTCTTCTGCTCGGTGTGTAAATGCTGCTTCTCCAGCAG 1959
 Db 2536 AsnGlnGluAlaArgThrIleThrAspArgIleGluArgIleGlnIleGlnTrpAsp 2555
 QY 1960 -----CTGTCTCAAC-----AAGAGCTCT 1977
 Db 2556 GluValGlnGluGlnLeuGlnAsnArgGlnGlnLeuAsnGluMetLeuLysAspSer 2575
 QY 1978 CTGAGTACCGAGCTCTGAGCAACAATTCACACAGCTGCTG----- 2019
 Db 2576 ThrGlnTrpLeuGluAlaLysGluGluAlaGlnValIleGlyGlnValArgGlyLys 2595
 QY 2020 -----GTTCCATATCTACGATCCAACTTACAGGA 2049
 Db 2596 LeuAspSerTrpLysGluGlyProHisThrValAspAlaIleGlnLysValThrGlu 2615
 QY 2050 TTTAAGCACTT-----GAAATCTTAATCTCAGAAACGGAGTGTGATGTG-----AAT 2100
 Db 2616 ThrLysGlnLeuAlaLysAspLeuArgGlnArgGlnIleSerValAspValAlaAsnAsp 2635
 QY 2101 AGCAAAATGAGTTGCTTGAAGTGTGCGCTGATTAACCTCACCAGGCTGTGACGGGTC 2160
 Db 2636 LeuAlaLeuLysLeuLeuArgAspTyrSerAlaAspThrArgLysValHisMetIle 2655
 QY 2161 TTGAGGATTTT-----CGTTCTGCT 2181
 Db 2656 ThrGluAsnIleAsnThrSerTrpGlyAsnIleHisLysArgValSerGluGlnGluAla 2675
 QY 2182 GCAGCCCCAACCACTCCATCTGGAATAGACAGCTACTCTCTCAACATTCACACAGCTGAT 2241
 Db 2676 AlaLeuGluThrHisArgLeuLeuGlnGlnPheProLeuAspLeuGlu- 2692
 QY 2242 CAGCAGATCCCTGAGGCTCTGCGCTGCGGCAATTTGCCCAATGTGTAAAGAAC--- 2298
 Db 2693 -----LysPheLeuSerTrpIleThrGluAlaGluThrThrAlaAsnVal 2707


```

QY 2299 -----GAGCGACTGAGGAGCGAGTGTGCG 2325
Db 2708 LeuGlnAspAlaSerArgIyGluLysLeuGluAspSerArgGlyValArgGluLeu 2727
QY 2326 TCGAACCCAGGATATGACAGCGAGGAGCTTGGAC-----GGTCTGGAACCAAGGC 2376
Db 2728 MetIyProTrpGlnAspLeuGlnGlyGluIleGluThrHisThrAspIleTyHisAsn 2747
QY 2377 CTCTGTGGCCCTGCGCACAAAGGATCGAGGTCTTCGCGAAGAGGAGCTCCATCGAGG 2436
Db 2748 LeuAspGluAsnGlyGlnLysIleLeuArgSerLeuGluGlySerAspGluAlaProLeu 2767
QY 2437 TTGCGAGATCACTCTGAAATCAAGCATACAAACTAGT-----GTTAAAGATTCCAA 2490
Db 2768 LeuGlnArgArgLeuAspAsnMetAnPhelysTrpSerGluLeuGlnLysIySerLeu 2787
QY 2491 AATCAACAAATACAGGTTAATCATGATAAAGAAATTTCTGAATTAATCTGACCGTAGAATAT 2550
Db 2788 AsnIleArgSerHisLeuGluAlaSerSerAspGlnTrpIyArgLeuHisLeuSerLeu 2807
QY 2551 GAAGAA 2556
Db 2808 GlnGlu 2809

RESULT 4
ID PGCV MOUSE STANDARD; PRT: 3358 AA.
AC Q62059; Q62058; Q9CUU0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN CSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
RC STRAIN=C57BL/6J; PubMed=7876137;
RX STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=9512551; PubMed=7822336;
RT Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RA "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT proteoglycan generated by alternative splicing.";
RL J. Biol. Chem. 270:958-965 (1995).
RN [2]
SEQUENCE FROM N.A. (ISOFORM V3).
RC STRAIN=C57BL/6J;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918 (1995).
RN [3]
SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tonaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,

```

```

RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tonita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilmink L.G., Wynshaw-Boris A., Yangisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki K., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
INTERACTION WITH FBLN1.
RP MEDLINE=99329059; PubMed=10400671;
RX Asperger A., Adam S., Kostka G., Timpl R., Heinigaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT versican.";
RL J. Biol. Chem. 274:20444-20449 (1999).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q62059-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
CC Name=V2;
CC IsoId=Q62059-3; Sequence=VSP_003089;
CC Name=V3;
CC IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
CC -!- TISSUE SPECIFICITY: V2 is found only in brain.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D16263; BAA03796.1; -.
DR EMBL; D28599; -. NOT ANNOTATED_CDS.
DR EMBL; D32040; BAA06802.1; -.
DR EMBL; AK014525; BAB29411.2; -.
DR HSSP; P01132; IEPG.
DR MGD; MGI:102889; Cagp2.
DR InterPro; IPR000152; Abx_hydroxyl_s.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.

```

[illegible]

Db 1684 rpValSerAspLysSerPheGluGlyValGlyLysGluLysGluAsnGluAspGluGluGlyA 1704
 QY 1208 CCATGAGGAGGACACACAGCAACAGAAATCTATCTCACAGCTACAGACCTCAAAAGGCTGA 1267
 Db 1704 laValAsnAlaAlaHisGlnGlyGluValArgAlaAlaThrGluArgSerAspHisLeuL 1724
 QY 1268 TCAGCAAGACTAGAGAGAAACAATCTTTGGATGTGGGGACAAATTCAGTTCAGTGATG 1327
 Db 1724 euLeuThrProGluLeuGluSerSerAsnValAlaSerSerAspLeuAlaThrTrpG 1744
 QY 1328 AA-----ATTGCTGGATCACTGCCAGCCTTTGGTCTCGACACCAATCAGACCTGCCCA 1381
 Db 1744 luGlyPheIleLeuGluThrThrProThr-----GluSerGluGlyGluMetAlaA 1761
 QY 1382 CATCTTTTGGCTTTATAAACAGAGGAT---GCTACTTTGATCCAGAACTTCCTCGTTGG 1438
 Db 1761 snSerThrProValPheArgGluThrIleGlyValAlaAsnValGluAlaGlnProPheG 1781
 QY 1439 AACCCGACTTGGACAGTGGACGGAGCAGAGCATGGTCTACCTGACACTTCTTGGTCTC 1498
 Db 1781 luHisSerSerSerHisProArgValGlnGluGluLeuThrThrLeuSerGlyAsnP 1801
 QY 1499 CACTGCTATG-----GCTCTACTCTCCCTGTCAGAGCTC 1534
 Db 1801 roProSerLeuPheThrAspLeuGlySerGlyAspAlaSerThrGlyMetGlu----- 1818
 QY 1535 CACCTTTCTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGACACACA----- 1587
 Db 1819 -----LeuIleThrAlaSerLeuPheThrIleuAspLeuGluSerGluThrLysValL 1836
 QY 1588 -----GATA 1591
 Db 1836 ysLysGluLeuProSerThrProSerProSerValGluIleSerSerSerPheGluPro 1856
 QY 1592 CAATGGCCACTGACCAAGATGCTA-----GTACAGGGCTCCACCATCCCAACCA 1642
 Db 1856 hrGlyLeuThrProSerThrValLeuAspIleGluIleAlaGlyValMetSerGlnThrS 1876
 QY 1643 GTGATTATTCGAATCAGCACTGGCTCTGGGANTTTCATCCATCCACTGATCTTCAG 1702
 Db 1876 erGlnLysThrLeuIleSerGlu-----IleSerGlyLysProThrSerGlnS 1892
 QY 1703 ATGACAGCGCATCAAGTGCAGGTGGGAGATATGGTCAGACACCTA----- 1749
 Db 1892 erGlyValArgAspLeuThrThrGlyPheProMetGlyGluAspPheSerGlyAspPheS 1912
 QY 1749 ----- 1749
 Db 1912 erGluThrProThrValSerThrProThrMetLysGluGluThrValGlyMetGlyGlyS 1932
 QY 1750 --GATCAAAATGATCTGTCTGACACTCTGCCCCATCTGAGTACAGAGCTCAGCGAAT 1807
 Db 1932 exAspAspGluArgValArgAspThrGlnThrSerSerIleProThrThrSerAspA 1952
 QY 1808 ATGTT---TCGTCCCGATCATTTCTTGGAGGATACCACTCTGCTCTCAGCTTTACAGT 1864
 Db 1952 snIleThrProValProAspSerLysGlyProAspSerThrValAlaSerThrThrAlap 1972
 QY 1865 ATATCACCAGTCTATGATGACCATGTCCTCCCAACAGCTCTTCAACAGAGCTCTCTGG 1921
 Db 1972 heProThrGluGluValMetSerSerAlaGluGlySerGlyGluGluLeuAlaSerVala 1992
 QY 1922 TCAGTCTGCTGTGTGCTAATGATGCTCTCCCAACAGCTCTTCAACAGAGCTCTCTGG 1981
 Db 1992 rgSerSerValGlyProValLeuProLeuAlaValAspIlePheSerGlyThrGluSerP 2012
 QY 1982 AGTACCGAGCTTGGAGCAACAATTCACAGCTG----- 2016
 Db 2012 roTyr-----PheAspGluGluPheGluGluValAlaAlaValThrGluAlaAsnGluA 2030
 QY 2017 -----CTGGTTCCATATCTAGCATCCATCTTACAGATTAAAGCACTTGAATAC 2068
 Db 2030 rgProThrValLeuProThrAlaAlaSerGlyAsnThr-----Vala 2044

QY 2069 TTAACCTCAGAAACGGAGTGTGATTTGTAATAGCAAAATG----- 2109
 Db 2044 epLeuThrGluAsnGlyTyrIleGluValAsnSerThrMetSerLeuAspPheProGlnT 2064
 QY 2110 -----AAGTTTCTAGTCTGTGCTCCCTATAACCTCACCAGAGCTGTGCACG 2155
 Db 2064 hrMetGluProSerLysLeuTrpSerLysProGluValAsnLeuAspLysGlnGluLeG 2084
 QY 2156 GG-----GTCCTTGAGAGATTTTCGTT 2176
 Db 2084 IyAzgGluThrValThrLysGluLysAlaGlnGlyGlnLysThrPheGluSerLeuHis 2104
 QY 2177 CTGCT---CCAGCCCAACACTCCATCTGGAATAGACAGTACTCTCTCAACATTGAAC 2233
 Db 2104 erSerPheAlaProGluGlnThrIleLeuGluThrGlnSerLeu-----IleGluT 2121
 QY 2234 CAGCTCATCAGCAGATCCCTGCAAGTTCTGCGCTGCGCGCGAATTTGCCAATGTGTAA 2293
 Db 2121 hrGluPheGlnThrSerAspTyrSerMetLeuThr-----ThrLeuLysThrTyrIleT 2139
 QY 2294 AGAACCAAGCAGCTGAGAGAGCGAGTGTGCTGCAACCCAGATATCAGACCGCGGA 2353
 Db 2139 hrAsnLysGluValGluGlu----- 2145
 QY 2354 GCTTGGACGCTCTG-----GAACACAGCCTCTGTGCGCTGCGCAACAGGAATGCG 2404
 Db 2146 -----GluGlyMetSerIleAlaHisMetSerThrProGlyProGlyIleLysAspLeuG 2164
 QY 2405 AG----- 2406
 Db 2164 IuSerThrThrHisProGluAlaProGlyLysSerHisSerPheSerAlaThrAlaL 2184
 QY 2407 --GTCTCCAGGAAAGGAGGATCCATGTCAGGTTGCCAGATCCTCTGAAATCAGCAT 2464
 Db 2184 euValThrGluSerGlyAlaAlaArgSerValLeuMetAspSerThrGlnGlu----- 2202
 QY 2465 ACACAACTACTGTTTAAAGTTCCAA-----AATCAACAAATAACA 2506
 Db 2203 --GluGluSerIleLysLeuPheGlnLysGlyValLysLeuThrAsnLysGluSerAsnA 2222
 QY 2507 AGGTAATCAGT 2517
 Db 2222 IaAspLeuSer 2225
 RESULT 5
 SNEI_HUMAN
 ID SNEI_HUMAN STANDARD; PRT; 8797 AA
 AC Q8NF91; Q84890; Q8N9P7; Q8TCPI; Q8WWM6; Q8XWF6; Q8WV17; Q8WV17;
 AC Q9COA7; Q9H525; Q9H526; Q9NS36; Q9NU50; Q9UU06; Q9UU07; Q9ULF8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic
 DE nuclear envelope protein 1) (Syn-1) (Myocyte nuclear envelope protein
 DE 1) (Wye-1) (Enapin)
 GN SYNE1 OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), CHARACTERIZATION, AND
 RP MUTAGENESIS OF 8758-LEU--CYS-8763.
 RC TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;
 RX MEDLINE=21652858; PubMed=11792814;
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
 RA Weisberg P.L., Ellis J.A., Shanahan C.M.;
 RT "Nesprins: a novel family of spectrin-repeat-containing proteins that
 RT localize to the nuclear membrane in multiple tissues";
 RL J. Cell Sci. 114:4485-4498(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.

CC TISSUE=Heart, Spleen, and Testis;
 RX MEDLINE=22296983; PubMed=12408964;
 RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
 RT "The neprins are giant actin-binding proteins, orthologous to
 RL Drosophila melanogaster muscle protein MSP-300.";
 RN Genomics 90:473-481(2002).
 RP [3]
 RP SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.
 RA Braune S., Abraham S., Padmakumar V., Tunggal B., Noegel A.A.,
 RA Korenbaum E.;
 RT "The longest isoform of enaptin/Syne-1, a nuclear envelope associated
 RT protein, binds actin cytoskeleton via the alpha-actinin-like actin-
 RT binding domain.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).
 RA Zhang Q., Shanahan C.M.;
 RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M.,
 RA Tracey A., Williams S.;
 RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 1-856 FROM N.A.
 CC TISSUE=Kidney;
 CC Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;
 RA "Golgi localization of syne-1.";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.
 CC TISSUE=Adrenal gland, and Teratocarcinoma;
 RA Niimiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Fuyuta T., Takahashi M., Kikkawa B., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otaki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Takemoto M., Ota T., Suzuki Y., Sugano S., Nagahara K., Masuno Y.,
 RA Nagai K., Isogai T.;
 RT "NED0 human cDNA sequencing project.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).
 CC TISSUE=Brain;
 CC MEDLINE=21082932; PubMed=11214970;
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:347-355(2000).
 RN [9]
 RP SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).
 CC TISSUE=Brain;
 CC Ansong W., Winkner U., Mewes H.-W., Weil B., Wiemann S.;
 RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).
 CC TISSUE=Brain;
 CC MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hiroseawa M., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:337-345(1999).
 RN [11]
 RP SEQUENCE OF 6922-8797 FROM N.A.
 CC TISSUE=Brain;
 CC MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:277-286(1998).
 RN [12]
 RP REVISIONS.
 CC TISSUE=Brain;
 CC MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106(2002).
 RN [13]
 RP SEQUENCE OF 7631-8797 FROM N.A., AND CHARACTERIZATION.
 CC MEDLINE=21659781; PubMed=11801724;
 RA Mislow J.M.K., Kim M.S., Davis D.B., McNally E.M.;
 RT "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner
 RT nuclear membrane, interacts with lamin A/C.";
 RL J. Cell Sci. 115:61-70(2002).
 RN [14]
 RP SEQUENCE OF 8406-8797 FROM N.A.
 CC Ma F.-R., Zhu L.-F.;
 RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Involved in the maintenance of nuclear organization and
 CC structural integrity. Probable anchoring protein which tethers the
 CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
 CC by interacting with the nuclear envelope and with F-actin in the
 CC cytoplasm.
 CC -!- SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal
 CC domain, and with LMNA in vitro (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
 CC largest part of the protein is cytoplasmic, while its C-terminal
 CC outer is associated with the nuclear envelope, most probably the
 CC nuclear nuclear membrane. In skeletal and smooth muscles, a
 CC significant amount is found in the sarcomeres.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=9;
 CC Name=1;
 CC IsoId=Q8NF91-1; Sequence=Displayed;
 CC Name=2; Synonyms=Beta;
 CC IsoId=Q8NF91-2; Sequence=VSP_007130;
 CC Name=3; Synonyms=Alpha;
 CC IsoId=Q8NF91-3; Sequence=VSP_007132, VSP_007144;
 CC Name=4;
 CC IsoId=Q8NF91-4; Sequence=VSP_007134, VSP_007139, VSP_007140,
 CC VSP_007144;
 CC Name=5;
 CC IsoId=Q8NF91-5; Sequence=VSP_007135, VSP_007136;
 CC Note=No experimental confirmation available;
 CC Name=6;
 CC IsoId=Q8NF91-6; Sequence=VSP_007137, VSP_007138;
 CC Note=No experimental confirmation available;
 CC Name=7;
 CC IsoId=Q8NF91-7; Sequence=VSP_007141, VSP_007142;
 CC Note=No experimental confirmation available;
 CC Name=8; Synonyms=Beta 2;
 CC IsoId=Q8NF91-8; Sequence=VSP_007131;
 CC Name=9; Synonyms=Alpha 2;
 CC IsoId=Q8NF91-9; Sequence=VSP_007133, VSP_007143, VSP_007144;
 CC TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal
 CC and smooth muscles, heart, spleen, and peripheral blood
 CC leukocytes.
 CC -!- DOMAIN: The Klarsicht domain, which contains a transmembrane
 CC domain, mediates the nuclear envelope targeting.
 CC -!- SIMILARITY: Belongs to the neprin family.
 CC -!- SIMILARITY: Contains 1 actin-binding domain.
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -!- SIMILARITY: Contains 12 HAT repeats.
 CC -!- SIMILARITY: Contains 1 Klarsicht domain.
 CC -!- SIMILARITY: Contains 31 spectrin repeats.
 CC -!- CAUTION: Ref.5 (CAB55865, CAB55866, CAC16280 and CAC16281)
 CC sequences differ from that shown due to erroneous gene model
 CC prediction.

-!- CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to a chimeric cDNA.
-!- CAUTION: Ref.14 sequence differs from that shown due to two framehifts in positions 8412 and 8784.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

ENBL; AY061755; AAL33798.1; -
ENBL; AY061756; AAL33799.1; -
ENBL; AF495910; AAN60442.1; -
ENBL; AF535142; AAN03486.1; -
ENBL; AY184203; AAO27771.1; -
ENBL; AY184205; AAO27774.1; -
ENBL; AL049548; CAB55865.1; -
ENBL; AL049548; CAB55866.1; -
ENBL; AL078582; CAB87586.1; -
ENBL; AL136079; -; NOT ANNOTATED CDS.
ENBL; AL138832; CAC16280.1; ALT_SEQ.
ENBL; AL138832; CAC16281.1; ALT_SEQ.
ENBL; AL357081; -; NOT ANNOTATED CDS.
ENBL; AL450401; -; NOT ANNOTATED CDS.
ENBL; AL589963; -; NOT ANNOTATED CDS.

```

ment Scores:

| | | | |
|-------------------|--------|---------------|------|
| No.: | 0.117 | Length: | 8797 |
| | 122.50 | Matches: | 173 |
| nt Similarity: | 38.04% | Conservative: | 145 |
| Local Similarity: | 20.69% | Mismatches: | 316 |
| Match: | 2.60% | Indels: | 202 |
| | 1 | Gaps: | 41 |

US-10-007-270-1 (1-3330) x SNE1 HUMAN (1-8797)

[illegible]

| | | | |
|----|------|--|------|
| QY | 1838 | ATACCACTCCTGCTCTCACGTTTACAGTATATACCACTATGTTCTCTATGACCATTCGCCCA | 1897 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: ::: | |
| Db | 6812 | -GlnSerAlalysAspArgLeuGluPheTrpThrGlnGlnSerValThrVal--- | 6830 |
| QY | 1898 | AGGCCCGAGACTGGTA-----GTGTTCTTC- | 1923 |
| Db | : | ::: ::: :: | |
| Db | 6830 | IngluLeuGluMetValArgAspHisLeuAsnAlaLpheLeuGluPheSerIysGluValA | 6850 |
| QY | 1924 | -----AGTCCGCTGTTTGCTAACTAGGCCTTCCAAAGACCTCTGTCACAAGA | 1972 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6850 | sPalaglInserSerLeuLysSerServaleuSerThrGlyAsnGlnLeuLeuArgLeuL | 6870 |
| QY | 1973 | GCTCTCTG-----GAGTACCGAGCTCTGGAGCAACAATTCAAC | 2011 |
| Db | : | ::: ::: :: | |
| Db | 6870 | yLysValAspThrAlaThrLeuArgSerGluLeuSerArgIleAspSerGlnTrpThra | 6890 |
| QY | 2012 | AGTCTCTG-----GTTCCATATCTACAGTCCAATCTTACAGGATTTACCAACTTGAAA | 2065 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6890 | sPLeuThrAsnIleProalaValGlnGluLysLeuHisGlnLeu---GlnMetAspL | 6909 |
| QY | 2066 | TACTTAACCTTCAGAACGGG----- | 2089 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6909 | yLeuProSerArgHisAlalleSerGluValMetSerTriPileSerLeuMetGluSenv | 6929 |
| QY | 2090 | TGATTGTGAATAGCAAATGAAGTTGCTAGTCTGTCGCGTATACCTACCAAGGCTG | 2149 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6929 | allieGlnLysAspGluAspAsnIleLysAsnSerileGlyTy------LysAlaI | 6946 |
| QY | 2150 | TCCAGGGGCTTGGAGGATTTTCGTTCTGCTGCAGCCCAACTCCATCTCGAATAAG | 2209 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6946 | IeHisGluTyLeuGlnLysTyLysGly-----PhelysIleAspIlea | 6961 |
| QY | 2210 | ACAGTACTCTCTCAACATTCGACCAGCTGATCAAGCAGATCCTCGAAGTTCTCGCCT | 2269 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6961 | enCyeLysGlnLeuthrValAspPheValAsnGlnSer-----ValL | 6975 |
| QY | 2270 | GGCGGGAATTTGCCOATGTCTAAAGAACGACCGACTGAGGAAGCGAGTGCCTGCA | 2329 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6975 | eugInIleSerSerGlnAspValGluSerLysArgSerAspLysThrAspPhea---- | 6993 |
| QY | 2330 | AACCAGGATATGACAGCCAGCGGAGCTCGAGCTGTGGAAACGAGCCTCTGTGGCCTG | 2389 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 6994 | -----GluInLeuGIyAlaMetAsnLysSerTrpGlnIleLeuGlnGlyLeuv | 7010 |
| QY | 2390 | GCACAAAGGAATGCAGGCTCTCCAGGAAAGGAGCTCCATCGAGGTGCCAGATCAT | 2449 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 7010 | alThrGluIuyelleGlnLeuLeuGluGly-----LeuudIuSerTrps | 7025 |
| QY | 2450 | CTGAAAAATCAAGCATACAAACTAGTGTAAA-----AAGTTCAAAAATCAAC | 2497 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 7025 | erGlu-----TyGIuAsnAsnValGlnCysLeuLysThrTrpPheIuThrGlnG | 7042 |
| QY | 2498 | AAATAACAAGGTAAATCAGTAAAAAGAAATCTCGAATTTACTGACCTGAGATATGAAGAT | 2557 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 7042 | LulyArgLeuLysGlnHIsarglleGlyAspGlnAlaSerValGlnAsnAlaLeuL | 7062 |
| QY | 2558 | TTAACCATCAGATTGGAGAGGAATTTAAAACTGAAA | 2595 |
| Db | : | ::: ::: :: ::: ::: ::: ::: ::: | |
| Db | 7062 | yspCysGlnAspLeuGluAspLeuIlleySalalys | 7074 |

RESULT 6

| ID | ACIN_HUMAN | STANDARD; | PRT; 1341 AA. |
|----|--|-----------|---------------|
| AD | ACIN HUMAN | STANDARD; | PRT; 1341 AA. |
| AC | Q9UKV3; O75158; Q9UG91; Q9UKV1; Q9UKV2; | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | |
| DE | Apoptotic chromatin condensation inducer in the nucleus (Acinus) | | |
| GN | ACINUS OR KRAA0670. | | |
| OS | Homo sapiens (Human) | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |

NCBI_TaxID=3606;
[1] SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), PARTIAL SEQUENCE, FUNCTION,
RN AND MUTAGENESIS OF ASP-1093
RX MEDLINE=99418558; Pubmed=10490026;
RA Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;
RT "Acinus is a caspase-3-activated protein required for apoptotic
RL chromatin condensation.";
RN Nature 401:168-173(1999).
[2] SEQUENCE FROM N.A. (ISOFORM 4).
RP TISSUE=Uterus;
RC Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3] SEQUENCE FROM N.A. (ISOFORM 3).
RP TISSUE=Fetal brain;
RC Li W.B., Gruber C., Jessee J., Polayes D.;
RA "Full-length cDNA libraries and normalization.";
RT Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4] SEQUENCE OF 56-1341 FROM N.A. (ISOFORM 1).
RP TISSUE=Brain;
RC MEDLINE=98403880; Pubmed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RN The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
CC -1- FUNCTION: Induces apoptotic chromatin condensation after
CC activation by CASP3.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=L;
CC IsoId=Q9UKV3-1; Sequence=Displayed;
CC Name=2; Synonyms=S';
CC IsoId=Q9UKV3-2; Sequence=VSP_004025; VSP_004028;
CC Name=3; Synonyms=S;
CC IsoId=Q9UKV3-3; Sequence=VSP_004026; VSP_004029;
CC Name=4;
CC IsoId=Q9UKV3-4; Sequence=VSP_004027;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- PTM: Undergoes proteolytic cleavage; the processed form is active,
CC contrary to the uncleaved form.
CC -1- SIMILARITY: Contains 1 SAP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.ebi.ac.uk/eb-
CC/announce/](http://www.ebi.ac.uk/eb-
CC/announce/) or send an email to license@ebi.ac.uk).

| | | |
|----------------------|---|---|
| EMBL; AF124726; | AA056724.1; | - |
| EMBL; AF124727; | AA056725.1; | - |
| EMBL; AF124728; | AA056726.1; | - |
| EMBL; AL050382; | CAB43681.1; | - |
| EMBL; BX247975; | CA062309.1; | - |
| EMBL; AB014570; | BAA31645.2; | - |
| Genew; HGNC:17066; | ACINUS. | |
| MMI; 604562; | - | |
| GO; GO:0005634; | C:nucleus; IDA. | |
| GO; GO:0016887; | F:ATPase activity; NAS. | |
| GO; GO:0019899; | F:enzyme binding; NAS. | |
| GO; GO:0003676; | F:nucleic acid binding; NAS. | |
| GO; GO:0030263; | P:apoptotic chromosome condensation; IDA. | |
| GO; GO:0030218; | P:apoptotic cell differentiation; IEP. | |
| GO; GO:0043657; | P:erythrocyte differentiation; IEP. | |
| InterPro; IPR003034; | SAP. | |
| Pfam; PF02037; | SAP; 1. | |

DR SMART; SM00513; SAP; 1.
 DR PROSITE; PS00800; SAP; 1.
 KW Apoptosis; Nuclear protein; Alternative splicing.
 FT DOMAIN 72 106 SAP.
 FT DOMAIN 142 442 GLU-RICH.
 FT DOMAIN 573 676 SER-RICH.
 FT DOMAIN 1114 1131 PRO-RICH.
 FT DOMAIN 1132 1341 ARG/ASP/GLU/LYS-RICH.
 FT SITE 1093 1094 CLEAVAGE (BY CASPASE-3).
 FT VARSPPLIC 1 727 Missing (in isoform 2).
 FT VARSPPLIC 1 758 /FTid=VSP_004025.
 FT VARSPPLIC 1 1152 Missing (in isoform 3).
 FT VARSPPLIC 1 1152 /FTid=VSP_004026.
 FT VARSPPLIC 728 766 /FTid=VSP_004027.
 FT VARSPPLIC 759 766 Missing (in isoform 4).
 FT VARSPPLIC 759 766 GSPKKCEAEAPPAQTQPTSTQTSHPESRIHVT
 FT P (in isoform 2). -> MSPADRCRSANTIEPATTSALFLLLQDQSSRTGL
 FT SERIHTTV -> MLSEKSG (in isoform 3).
 FT D-2A: ABOLISHES CLEAVAGE BY CASP3 AND
 FT CHROMATIN CONDENSATION ACTIVITY.
 FT Q -> H (in Ref. 4).
 SQ SEQUENCE 1341 AA; 151887 MW; 8FE286681F83AB5C CRC64;

Alignment Scores:
 Pred. No.: 0.083 Length: 1341
 Score: 151.00 Matches: 1341
 Percent Similarity: 33.21% Conservative: 157
 Best Local Similarity: 19.75% Mismatches: 107
 Query Match: 2.57% Indels: 301
 DB: 1 Gaps: 32

US-10-007-270-1 (1-3330) x ACIN_HUMAN (1-1341)

QY 314 GATTTCGCAAGCCTCGAACAAGATCCCAATTTTCCCAAG-----GGCGTT 364
 Db 110 AsnLeuGlnLysHisThrProHisAlaAlaPheGlnProAsnSerGlnIleGlyGlu 129
 QY 365 AAAGTCGTGCCACAGGAATCCATGAACAGATTTAGACAGCTCTTCAGCTTATTATAGA 424
 Db 130 GluSerGlnAsnSerPheIleLysGlnTyLeuGlnLysGlnGlnLeuLeuArg 149
 QY 425 TTGAGATGTGTACGAGAACGATATGGGAAGCATATCGGATCTTTCTGGATCGCATCCCT 484
 Db 150 GlnArgLeuGlnArgGluAla-----Arg 157
 QY 485 GACACAGGGGATATACAGCTGGCTCAGCATCTGCACAGAGAGACCTTCGCTCTTT 544
 Db 158 GluAlaAlaGluLeuGluGlu-----AlaSerAlaGluSerGluAspGluMetIleHis 175
 QY 545 GACATTGGA-----AAAACTTCAGCAATTTCCAGAGAGACCTCGAT 586
 Db 176 ProGluGlyValAlaSerLeuLeuProProAspPheGlnSerSerLeuGlnArgProGlu 195
 QY 587 CTTCTCCAGCAGAGATAAACAAGAGAAGTTTCCTTCAGACAAAAGATCAATCTCGCA 646
 Db 196 Leu-----GluLeuSerArgHisSerProArgLysSerSerIleSerGlu 211
 QY 647 GAGAGACATTGGGAGAGCTGGTGAACCATTTGCTATTTCAACAGCAATCTCATTTCA 706
 Db 212 GluLysGlyAspSerAspAsp-GluLysProArgLysGlyGluArgArgSerSerArgVa 231
 QY 707 AAGA-----CTTGGCAGATTTCTTAAGAAACCTTCACAGAGCAA 748
 Db 231 lArgGlnAlaArgAlaAlaLysLeuSerGluGlySerGlnProAla-GluGlnGluAla 251
 QY 749 TTCAGATGTTGCCAAGCTCTCACTTGGGCCCTTCCCTCTCACTCTCTGATGACACCTCC 808
 Db 251 spGlnGlu-----ThrProSerArgAsnLeu- 259
 QY 809 TCAATGAAATTCGATATACACTCAACGACACCAAGATGCTCTACAGAAAGAGAAA 868

Db 536 euAlaLeuAlaLysGlyIleThrGluGluCysLeuLysGlnProSerLeuGluGlnLysG 556
 QY 1801 -----AGCAATATGTTCTGTCAGATCATTTCTTGAGGATACCACTCTG 1849
 Db 556 luGlyArgAlaSerHisThrLeuLeuProSerHisArgLeuLysGlnSerAlaAsp 576
 QY 1850 TCTCAGCTTTACAGTATATACCACTAGTCTTATGACATTCGCCCCCAAGCGCGAGAC 1909
 Db 576 erSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerArg 594
 QY 1910 TGCTAGTGTCTTCAGTCTGCTGTGTCTTAATGCGCTTCTCCCAACGCTGTTCAACA 1969
 Db 595 -----SerProAsp-----SerS 599
 QY 1970 AGAGCTCTCTGAGTACGAGCTCTGGAGCAACAATTCACAGCTGCTGCTGCATATC 2029
 Db 599 erGlySerArgSerHisSerProLeuArgSerLysGlnArgAspValAlaGlnAlaArgT 619
 QY 2030 TAGCATCCCACTTACAGGATTTAAGCACTTGAATATCTTAATCTCAGAACCGGAGTG 2089
 Db 619 hrHisAlaAspProArgGlyArgProLys-----MetGlySerArgSerThrSerG 636
 QY 2090 TGATGTGATGACAAATGAAGTTTCTGAAGTCTGCGGTATACCTCACCAGGCTG 2149
 Db 636 luSerArgSerArgSerArgSerArgSerArgSerArgSerArgSerArgLysSerL 656
 QY 2150 TGACAC---GGGGTCTGGAGATTTCTGCTGCTGCGACCCCAACTCCATCTCGAAA 2206
 Db 656 euSerProGlyValSerArgSerSerThr----- 666
 QY 2207 TAGACAGCTACTCTCTCAACATTTGAACAGCT-----GATCAGCAGATCTCTCA 2257
 Db 667 -----SerThrThrGluThrLysAspProSerSerGlyGlnGluValAlaThrPro 683
 QY 2258 AGTTCTGCTGGCGGAAATTTGCCAATCTGTAAGAACGAAACGAGCTAG----- 2310
 Db 684 -----ProValProGlnLeuGlnValCysGluProLysGluArgThrSerThrSer 701
 QY 2311 -----GAAGCGAGTCTGCTGCAACACGAGTATGACAGCCAGGAGCTCTGG 2359
 Db 701 erSerSerValGlnAlaArgLeuSerGlnProGluSerAlaGluLysHisValThrG 721
 QY 2360 AGGFTCTGGAACAGGCTCTGTCGCTGGCACAAGGATGCGAGTCTCTCCAGGAA 2419
 Db 721 InArgLeuGlnProGluArgGlySerPro-----LysLysCysGluAlaGluGluAlaG 739
 QY 2420 AGGAGCTCCATGCGAGTGTGCGAGTCACTCTGAAATCAA 2460
 Db 739 luProProAlaAlaThrGlnProGlnThrSerGluThrGln 752
 RESULT 7
 ID RTN4 RAT
 AC Q9JKL1; Q9JKL1; Q9R009; Q9WUE9; Q9WUFO; PRT; 1163 AA.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 10-OCT-2003 (Rel. 41, Last sequence update)
 DE Reticulon 4 (Neurite outgrowth inhibitor) (Nogo protein) (Foccen)
 DE (Glut4 vesicle 20 kDa protein).
 GN RTN4 OR NOGO.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND PARTIAL SEQUENCE.
 RC STRAIN-Sprague-Dawley; TISSUE=Adipocyte;
 RX MEDLINE=9249816; PubMed=10231557;
 RA Morris N.J., Ross S.A., Neveu J.M., Lane W.S., Lienhard G.E.;
 RT "Cloning and characterization of a 22 kDa protein from rat adipocytes:
 RL a new member of the reticulon family.";
 RL Biochim. Biophys. Acta 1450:68-76(1999).
 RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RX MEDLINE=20129258; PubMed=10667796;
 RA Chen M.S., Huber A.B., Van der Haar M.E., Frank M., Schnell L.,
 RA Spillmann A.A., Christ F., Schwab M.E.;
 RT "Nogo-A is a myelin-associated neurite outgrowth inhibitor and an
 antigen for monoclonal antibody IN-1.";
 RL Nature 403:434-439(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
 RC STRAIN-Wistar Kyoto; TISSUE=Vascular smooth muscle;
 RA Ito T., Schwartz S.M.;
 RT "Cloning of a member of the reticulon gene family in rat: one of two
 RT minor splice variants.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP FUNCTION.
 RX MEDLINE=22033691; PubMed=12037567;
 RA GrandPre T., Li S., Strittmatter S.M.;
 RT "Nogo-66 receptor antagonist peptide promotes axonal regeneration.";
 RL Nature 417:547-551(2002).
 CC -!- FUNCTION: Potent neurite outgrowth inhibitor which may also help
 CC block the regeneration of the nervous central system in adults (By
 CC similarity).
 CC -!- SUBUNIT: Binds to RTN4R. Interacts with Bcl-xL and Bcl-2 (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Anchored to the
 CC membrane of the endoplasmic reticulum through 2 putative
 CC transmembrane domains (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=1; Synonyms=Nogo-A, NI-220-250;
 CC IsoId=Q9JKL1-1; Sequence=Displayed;
 CC Name=2; Synonyms=Nogo-B, Foccen-M1;
 CC IsoId=Q9JKL1-2; Sequence=VSP_005658;
 CC Name=3; Synonyms=Nogo-C, VP20;
 CC IsoId=Q9JKL1-3; Sequence=VSP_005656, VSP_005657;
 CC Name=4; Synonyms=Foccen-M2;
 CC IsoId=Q9JKL1-4; Sequence=VSP_005659;
 CC -!- TISSUE SPECIFICITY: Isoforms 1, 2 and 3 are present in optic
 CC nerve, spinal cord and cerebral cortex. Isoforms 1 and 2 are
 CC present in dorsal root ganglion, sciatic nerve and PC12 cells
 CC after longer exposure. Isoforms 2 and 3 are detected in kidney,
 CC cartilage, skin, lung and spleen. Isoform 3 is expressed at high
 CC level in skeletal muscle. In adult animals isoform 1 is expressed
 CC mainly in the nervous system.
 CC -!- SIMILARITY: Contains 1 reticulon domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF051335; AAF01564.1; -
 DR EMBL; AJ242961; CAB71027.1; -
 DR EMBL; AJ242962; CAB71028.1; -
 DR EMBL; AJ242963; CAB71029.1; -
 DR EMBL; AF132045; AAD31019.1; -
 DR EMBL; AF132046; AAD31020.1; -
 DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IDA.
 DR GO; GO:0005635; C:nuclear membrane; ISS.
 DR GO; GO:0005515; F:protein binding; ISS.
 DR GO; GO:0019987; F:negative regulation of anti-apoptosis; ISS.
 DR GO; GO:0030517; F:negative regulation of axon extension; ISS.
 DR InterPro; IPR003388; Reticulon.
 DR Pfam; PF02453; Reticulon; 1.
 DR PROSITE; PS0845; RETICULON; 1.
 DR Endoplasmic reticulum: Alternative splicing; Transmembrane.
 DR DOMAIN 1 989
 DR TRANSMEM 990 1010
 DR DOMAIN 1011 1104
 DR LUMENAL (Potential).


```

Db 723 pLeuPheSerAspSerIleProGluValProGlnThrGlnGluAlaVal----- 741
QY 1923 CAGTCTCGGTGTCTAACTGCTTCCCAACGACTGTTCCACAGAGCTCTTCGGA 1982
Db 742 -----MetLeuMetIysGluSerLeuThrGluValSerGluThrVa 755
QY 1983 GTACCGAGCTCTGGAGCAACAATTCACA-----CAGCTGCTGGTTCATATCT 2030
Db 755 lAlaGlnHisGluGluArgLeuSerAlaSerProGlnGluLeuGlyIysProTyrIle 775
QY 2031 ACGATCC-----AATCTACAGATTAAAG----- 2055
Db 775 uGluSerPheGlnProAsnLeuHisSerThrIysAspAlaAlaSerAsnAspIleProTh 795
QY 2056 -----CAACTTGAATTAATTAAC-----TTGAG 2078
Db 795 rLeuThrIysIysGluIysIleSerLeuGlnMetGluGluPheAsnThrAlaIleYrSe 815
QY 2079 AAACGGAGGTGTAATGTGAATAGCAAAATGAAGTTTGCTAAGTCTGCGGTATAACT 2138
Db 815 rAsnAspAspLeuLeuSerSerIysGluAspIleIleIysGluSerGluThrPheSerAs 835
QY 2139 CACCAAGCTGTGCACGGGTCTTGAGGATTGTTCT----- 2178
Db 835 pSerSerProIleGlu-----IleIleAspGluPheProThrPheValSerAlaIysAspAs 854
QY 2179 -----GCTGCAGCCCAACCACTCCATCTGGAATAGACAGCTACTCT-----CT 2222
Db 854 pSerProIysLeuAlaIysGluTyrThrAspLeuGluValSerAspIysSerGluIleAl 874
QY 2223 CACATTGACAGCTGATCAAGCAGATCCCTGCAAGTTCTCTGCGCTGCGGC---GAATT 2279
Db 874 aAsnIleGlnSerGlyAlaAspSerLeuProCysLeuGluLeuProCysAspIleSerPh 894
QY 2280 TGCCCAATGT-----GTAAAGAACGAACGAGTCAAGAGCGGA 2318
Db 894 eIysAsnIleTyrProIysAspGluValHisValSerAspGluPheSerGluAsnAspSe 914
QY 2319 GTGTGCTGCAACACGAGATATGACAGCGAGGCTGCGAGCTGCGACGAGTCAAGAGCG 2378
Db 914 rSerValSerIyAlaSerIle---SerProSerAsnValSerAlaLeuGlu----- 930
QY 2379 CTGTGGCTGCGCAACAGGAATCGAGGTCTCCAGGGAAGGGA-----GC 2426
Db 931 -----ProGlnThrGluMetGlySerIleValIysSerIysSerLeuThrIysGluAl 948
QY 2427 TCCATGCAAGTGTCCAGATCACTCTGGAATCA 2460
Db 948 aGluIysLeuProSerAspThrGluIysGlu 959
RESULT 8
BPAL_MOUSE
ID BPAL_MOUSE STANDARD; PRT; 7389 AA.
AC Q91ZU6; Q91ZU7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bulbos pemphigoid antigen 1, isoforms 1/2/3/4 (BPA) (Hemidesmosomal plaque protein) (Dystonia musculorum protein) (Dystonin).
GN BPAG1 OR DST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RX STRAIN=BALE/c; TISSUE=Muscle, and Neuron;
RX MEDLINE=21405767; PubMed=11514586;
RA Leung C.L., Zheng M., Prater S.M., Liem R.K.H.;
RT "The BPAG1 locus: alternative splicing produces multiple isoforms with distinct cytoskeletal linker domains, including predominant isoforms in neurons and muscles.";
RL J. Cell Biol. 154:691-697 (2001).

```

```

RN [2]
RP SEQUENCE OF 6693-7389 FROM N.A. (ISOFORMS 3 AND 4).
RC STRAIN=C57BL/6J; TISSUE=Fetal skin, and Fetal spinal cord;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Mateuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmer S., Gutincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Walstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
CC -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC intermediate filaments to the inner plaque of hemidesmosomes. The
CC proteins may self-aggregate to form filaments or a two-dimensional
CC mesh (By similarity).
CC -!- SUBUNIT: Homodimer. Interacts with the neuronal intermediate
CC filament protein, Prph (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Name=2; Synonyms=b;
CC IsoId=Q91ZU6-1; Sequence=Displayed;
CC Name=1; Synonyms=a;
CC IsoId=Q91ZU6-2; Sequence=VSP_050483;
CC Name=3;
CC IsoId=Q91ZU6-3; Sequence=VSP_050484, VSP_050485, VSP_050486;
CC Note=No experimental confirmation available;
CC Name=4;
CC IsoId=Q91ZU6-4; Sequence=VSP_050485, VSP_050486;
CC Note=No experimental confirmation available;
CC Name=5; Synonyms=e;
CC IsoId=Q91ZU8-1; Sequence=External;
CC Name=6; Synonyms=nl;
CC IsoId=Q60824-1; Sequence=External;
CC Name=7; Synonyms=n2;
CC IsoId=Q60824-2; Sequence=External;
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the heart and
CC skeletal muscle and at low levels in the skin in the adult.
CC Expressed in the myocardium, skeletal muscle masses, vertebrae
CC cartilage, and epithelia of the tongue of 14.5 day embryos.
CC -!- SIMILARITY: Belongs to the plakophilin or cytolinker family.
CC -!- SIMILARITY: Contains 1 actin-binding domain.
CC -!- SIMILARITY: Contains 1 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- SIMILARITY: Contains 9 plectin repeats.
CC -!- SIMILARITY: Contains 27 spectrin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

[illegible]

EMBL; AF396879; AAK83384.1; -
EMBL; AF396878; AAK83383.1; -
EMBL; AK051626; BAC34695.1; -
EMBL; AK037206; BAC29753.1; -
MGD; MGI.104627; Dst.
GO; GO:0005737; C:cytoplasm; ISS.
GO; GO:0030056; C:hemidesmosome; IDA.
GO; GO:0052200; F:structural constituent of cytoskeleton; ISS.
GO; GO:0045104; P:intermediate filament cytoskeleton organiza. .; ISS.
InterPro; IPR001589; Actbind actnin.
InterPro; IPR001715; Calponin-like.
InterPro; IPR002048; EF-hand.
InterPro; IPR003108; GAS2.
InterPro; IPR001101; Plectin_repeat.
InterPro; IPR001452; SH3.
InterPro; IPR002017; Spectrin.
Pfam; PF00307; CH; 2.
Pfam; PF00306; ehand; 2.
Pfam; PF02187; GAS2; 1.
Pfam; PF00681; Plectin; 5.
Pfam; PF00016; SH3; 1.
Pfam; PF00435; spectrin; 26.
ProDom; PD000012; EF-hand; 1.
SMART; SM00033; CH; 2.
SMART; SM00054; Eph; 2.
SMART; SM0243; GAS2; 1.
SMART; SM00250; PLEC; 9.
SMART; SM00150; SPEC; 34.
PROSITE; PS00019; ACTININ_1; 1.
PROSITE; PS00021; CH; 2.
PROSITE; PS00018; EF_HAND; 2.
Actin-binding; Coiled coil; Repeat; SH3 domain; Structural protein;
Cytoskeleton; Cell adhesion; Calcium; Calcium-binding;
Alternative splicing.
DOMAIN 31 255 ACTIN-BINDING.
DOMAIN 35 138 CH 1.
DOMAIN 151 252 CH 2.
REPEAT 590 667 SPECTRIN 1.
REPEAT 675 770 SPECTRIN 2.
DOMAIN 889 941 SH3
REPEAT 1260 1356 SPECTRIN 3.
REPEAT 1537 1581 PLECTIN 1.
REPEAT 1582 1619 PLECTIN 2.
REPEAT 1657 1694 PLECTIN 3.
REPEAT 1695 1732 PLECTIN 4.
REPEAT 1735 1770 PLECTIN 5.
REPEAT 1771 1808 PLECTIN 6.
REPEAT 1811 1846 PLECTIN 7.
REPEAT 1847 1884 PLECTIN 8.
REPEAT 1886 1922 PLECTIN 9.
REPEAT 3814 3914 SPECTRIN 4.
REPEAT 4053 4152 SPECTRIN 5.
REPEAT 4270 4346 SPECTRIN 6.
REPEAT 4409 4517 SPECTRIN 7.
REPEAT 4522 4620 SPECTRIN 8.
REPEAT 4623 4729 SPECTRIN 9.
REPEAT 4742 4840 SPECTRIN 10.
REPEAT 4851 4949 SPECTRIN 11.
REPEAT 5177 5278 SPECTRIN 12.
REPEAT 5288 5385 SPECTRIN 13.
REPEAT 5397 5497 SPECTRIN 14.
REPEAT 5506 5605 SPECTRIN 15.
REPEAT 5646 5714 SPECTRIN 16.
REPEAT 5725 5824 SPECTRIN 17.
REPEAT 5946 6046 SPECTRIN 18.
REPEAT 6055 6155 SPECTRIN 19.
REPEAT 6165 6265 SPECTRIN 20.
REPEAT 6274 6372 SPECTRIN 21.

| | | | | |
|----|---------|------|------|--------------------------|
| FT | REPEAT | 6383 | 6480 | SPECTRIN 22. |
| FT | REPEAT | 6492 | 6592 | SPECTRIN 23. |
| FT | REPEAT | 6602 | 6701 | SPECTRIN 24. |
| FT | REPEAT | 6710 | 6808 | SPECTRIN 25. |
| FT | REPEAT | 6826 | 6914 | SPECTRIN 26. |
| FT | REPEAT | 6962 | 7020 | SPECTRIN 27. |
| FT | CA BIND | 7028 | 7040 | EF-HAND 1 (POTENTIAL). |
| FT | CA BIND | 7064 | 7076 | EF-HAND 2 (POTENTIAL). |
| FT | DOMAIN | 452 | 486 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 638 | 703 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 737 | 773 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 1003 | 1138 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 1195 | 1247 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 1413 | 1455 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 1504 | 1527 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 3336 | 3359 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 3539 | 3715 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 3809 | 3893 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 3957 | 3978 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 4006 | 4043 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 4159 | 4209 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 4270 | 4311 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 4426 | 4563 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 4596 | 4735 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 4847 | 5097 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 5173 | 5233 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 5570 | 5603 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 5717 | 5739 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 5787 | 5809 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 6010 | 6089 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 6116 | 6164 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 6277 | 6318 | COILED COIL (POTENTIAL). |
| FT | DOMAIN | 6381 | 6417 | COILED COIL (POTENTIAL). |

Alignment Scores:

| | | | |
|------------------------|--------|---------------|------|
| Pred. No.: | 0.184 | Length: | 7389 |
| Score: | 149.00 | Matches: | 214 |
| Percent Similarity: | 34.25% | Conservative: | 158 |
| Best Local Similarity: | 19.71% | Mismatches: | 387 |
| Query Match: | 2.54% | Indels: | 327 |
| DB: | 1 | Gaps: | 51 |

US-10-007-270-1 (1-3330) x BPA1 MOUSE (1-7389)

| | | |
|------|--|------|
| 170 | ATTTTCTCCAGGTTCAAGCAACAAAGATATCTCCATTAAACATATACCATCTTCGAACT | 229 |
| Db | | |
| 4839 | LeuLeuLeuYsThrGlnGluYsAlaIaLeuGlnLeuLeuAsnThrMet | 4858 |
| QY | | |
| 230 | AAAGACATAGACAATCCCCAGAAAGTAAACAACACTGAAAGTACTCAAAAAATG- | 283 |
| Db | | |
| 4859 | LysThrAspTrpAspArgPheArgGlyGlnValYsGluArgGluGluYsLeuYsAsp | 4878 |
| QY | | |
| 284 | -----TCAAAATGTCAACTATGACGACGAATATTCGAT | 316 |
| Db | | |
| 4879 | SerLeuGlnLysAlaLeuLysTyArgGluGlnValGluThrLeuArgProTrpIleAsp | 4998 |
| QY | | |
| 317 | TTGGCAAGACGATCGAACAAAAAGATCCGCAATTTTCCCAACGGGGTTAAAGTGTGTCCA | 376 |
| Db | | |
| 4899 | ArgCysGlnHisSerLeuAspGlyValThrPhe-----SerLeuAspPro | 4913 |
| QY | | |
| 377 | CAGGAATCCCATGAAACAGATT-----TTAGACAGTCTTCA- | 412 |
| Db | | |
| 4914 | ThrGluSerGlnSerSerIleAlaGluLeuYsSerLeuGlnLysGluMetAspHisHis | 4933 |
| QY | | |
| 413 | -----GCATTATTAGATTGACAGGTGCTCAG- | 439 |
| Db | | |
| 4934 | PheGlyMetLeuGluLeuLeuAsnAsnThrAlaAsnSerLeuLeuSerValCysGluVal | 4953 |
| QY | | |
| 440 | -----GAGCAGATATGGAGACATATCCGATCTTTCTGGATCGCATCCCTGCACACAGGG | 493 |
| Db | | |
| 4954 | AspYsGluAlaValThrGluGluAsnGlnSerLeuMetGluLysValAsnArgValThr | 4973 |
| QY | | |
| 494 | GAATATCAGGCTGGGTGACGATCTGCCAGCAGGAGACCTTCTGCCTCTTTGCACATTGGA | 553 |

4974 GluGln-----LeuGlnSerLysThrValSerLeuGluAsnMetal 4987
554 AAAAATCTTCAGCAATTCCTCCAGGAGCACCTGGATCTCTCCAGCAGAGATAAAACAGAGA 613
4988 GlnLysPheLysGluPheGlnGlu-----ValSerArgAspThrGlnLys 5002
614 AGTTCCCTCGACAGAAAGATGAATATCTCAGAGAAGACATTCGGAGAGCCCTGGTGA 673
5003 GlnLeuGlnAspThrLysGluGlnGluValHisSerLeuGlyProGlnAlaTyr 5022
674 -----ACCATGTCATTTCAACAGCAATCTCATTTCAAAAGATTCGGGCA 718
5023 SerAsnLysHisLeuSerValLeuGlnAla-GlnGlnLysSerLeuGlnThrLeuLysGln 5042
719 GTAT-----TCTAAGAAAACCTCAGAAAGAGCAAAATTCAGATGTGGCAACGCTCT 769
5042 nGlnValAspGluAlaLysArgLeu-AlaGlnAspLeuValGluAlaAlaAsp--S 5061
770 CACTTGGGCTTCCCTCTCATCTCTGTATGACACCTCTCTCAATGAAATTCGATATA 829
5061 erLysGly-----ThrSerAspValLeuLeuGlnAlaGluThrLeu----- 5074
830 CACTCAACGACACCAAGAGCCCTACACAGAAGAGAAACAGAAATTCGCTGTGTGGAGG 889
5075 -----AlaGluGluHisSerGluLeuSer----- 5082
890 AGCAGAGGTGGAGCTCAGCGTCTCTCTGTGTAACACAGAAG-----T 931
5083 --GlnGlnValAspGluLysCysSerPheLeuGluThrLysLeuGlnGlyLeuGlyHisP 5102
932 TCAAGGCAGAGCTCGCTGATCCCGAGTCCCGATATATACCAG-----GAGCTAGCAG 982
5102 heGlnAsnThrIleArgGluMetPheSerGlnPheThrGluCysAspAspGluLeuAspG 5122
983 GA-----AAGTCCCAACTTCAGATCGCAA 1006
5122 LysMetAlaProValGlyArgAspAlaGluThrLeuArgLysGlnLysAlaCysMetGlnT 5142
1007 AGATATTTAAGAACTTCAGGATTCAAAAAATCCATGTGTTAGNTTTAGACCAAGA 1066
5142 hrPheLeuLysLysLeuGluAlaLeuMetAlaSerAsnAspSerAlaAsnArgThrCysL 5162
1067 AAGAAAAGATGCTCAAGTCCACAGAGATCAACTTACGCCATCTTTAAGAGACACA 1126
5162 yMetMetLeuAlaThrGluGluThrSerProAspLeuIleGlyVal-----LysArgA 5180
1127 GTGCAGAGCAAAAGCCCTGCAAGTGACCTCTGCTTTGATTCACCAAAAATGAAA 1186
5180 spLeuGluAlaLeuSerLysGlnCysAsnLysLeu--LeuAspArgAlaLysThrArgG 5199
1187 GTGAGGAGTCTATCATGGAACCATGAGGAGGAGCAAGCAACCAAGATCTATCTACAG 1246
5199 luGluGlnValAspGlyAlaThr-----GluLysLeuGluGluPheHisArgLysL 5216
1247 CTACAGACCTCAAAAGCGTGATCAGCAAGCACTAGAGGAGCAACAATCTTTC--GATG 1303
5216 euGluGluPheSerThrLeuLeuGlnLysAlaGluGluHisGluLysSerGlnGlyProV 5236
1304 TGGGAGCAATTCAGTTCACTGATGAATTCGTGATCATCTGCCAGCCCTTGGTCTCGACA 1363
5236 alGlyThr-----GluThrGluT 5242
1364 CCCAMTACAGCTGCCACATCTTTTGCTGTTATACAGAGGATGCTACTTTGAGTCCAG 1423
5242 hrIleAsnGlnGlnLeuAspValPheLysValPheGlnLysGlu-----G 5257
1424 AACTTCTCTCTGTTGAACCCAGCTTGAGACAGTGGAC----- 1461
5257 luIleGluProLeuGlnValLysGlnGlnAspValAsnTrpLeuGlnGlyLeuIleG 5277
1462 -----GGACGAGCATGGTCTTACCT-----GACA 1486

| | | | |
|------------|--|--|--------------|
| QY | 2324 | GCTGCAAAACAGGATATGACAGCCAGGAGCGCTGGACGCTCTGCACACAGGCTCTGTG | 2383 |
| Db | 5610 | rgalalysGluGluLeuGlu---LysThrIleAlaAspAspAenGluPro----- | 5625 |
| QY | 2384 | GCCTGGCACAAGAAAGATGGAGGTCCTCCAGGGAAGGAGCTCCATGCGAGGTGGCCAG | 2443 |
| Db | 5626 | ---LeuProAspCysGlu-----Prot | 5632 |
| QY | 2444 | ATCACTCTGMAAATCAAGCATACAAACTACTGTTTAAAGTTCCAAATACACAAATA | 2503 |
| Db | 5632 | hrGlnSerArgHisLysValGluGluIleAspAlaIleLeuArgSerGlnPheG | 5652 |
| QY | 2504 | ACAAGTAAATCAGTAAAGAAATCTCGAATTAATCTGACCGTAGATATGAGAATTAAAC | 2563 |
| Db | 5652 | luGlnAla-----AlaAspAlaGluLeu----- | 5659 |
| QY | 2564 | ATCAAGATTGGGAAGGAATTTAAAACTGAATGTTACAAATTATCACTTAGCTATCTCA | 2623 |
| Db | 5660 | -----SerTrpIleThrGluThrGlnLysLysLeuMetSerLeuGlyAspIleArgLeuG | 5678 |
| QY | 2624 | AGAGAGATGATTGGCTTCTCAAGGAAATGGAGACAGCATATTCATGGGTGATC--- | 2679 |
| Db | 5678 | luGlnAspGlnThrSerAlaGlnLeuGlnValGlnLysAlaPheThrMetAspIleLeuA | 5698 |
| QY | 2680 | -----AAAAATCCAGACATACATC | 2698 |
| Db | 5698 | rgHisLysAspIleIleAspGluLeuValThrSerGlyHisLysIleMetThr---ThrSer | 5717 |
| QY | 2699 | AAACTGAGAAATCAGCACACCATATTTCAATATAGACAGTCAATG-----TACTTG | 2752 |
| Db | 5718 | GlyGluGluGluLysGlnSerMetLysLysLysLysLeuAspLysValLeuLysLysTy | 5737 |
| QY | 2753 | GCACCACTAAATTCGAAAAAAGACACACTTACTTATTATTAACCCCAATGCATC | 2812 |
| Db | 5738 | AlaValCysGlnIleAsnSerGluArgHisLeuGlnLeuGluArgAlaGlnSerLeuVal | 5757 |
| QY | 2813 | AGC-----GAAACATAT-----TTTTACTATCTTGGATGATA---GTCAAAATGATC | 2857 |
| Db | 5758 | SerGlnPheTrpGluThrTrpGluGluLeuLeuTrpProTrpLeuThrGluThrGlnArg | 5777 |
| QY | 2858 | ATAAGCCAGGTT | 2869 |
| Db | 5778 | IleSerGlnLeu | 5781 |
| RESULT 9 | | | |
| CRK7_HUMAN | | | |
| ID | CRK7_HUMAN | STANDARD; | PT; 1490 AA. |
| AC | Q9NVY4; O94978; | | |
| DT | 16-OCT-2001 (Rel. 40, Created) | | |
| DT | 16-OCT-2001 (Rel. 40, Last sequence update) | | |
| DT | 15-MAR-2004 (Rel. 43, Last annotation update) | | |
| DE | Cell division cycle 2-related protein kinase 7 (EC 2.7.1.-) (CDC2- | | |
| DE | related protein kinase 7) (Crkrs). | | |
| GN | CRK7 OR KIAA0904. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| ON | NCBI_TaxID=9606; | | |
| OX | (1) | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | MEDLINE=21539573; PubMed=11683387; | | |
| RX | Ko T.K., Kelly E., Pines J., | | |
| RT | "Crkrs; a novel conserved Cdc2-related protein kinase that colocalises | | |
| RT | with SC35 speckles"; | | |
| RL | J. Cell Sci. 114:2591-2603(2001). | | |
| RL | (2) | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=Brain; | | |
| RX | MEDLINE=22158633; PubMed=12168954; | | |
| RA | Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.; | | |
| RT | "Construction of expression-ready cDNA clones for KIAA genes: manual | | |
| RT | curation of 330 KIAA cDNA clones."; | | |

Db 290 -----ArgSerProSerProTyArgSer----- 298
 QY 657 TGGGAGAGCTGGTGAACCAATTCATTTCAACAGCAATCTACATTTCAAGACTGGG 716
 Db 299 ---GlnArgSerValSerProTyArgSerProTyArgSerSerSerTyGluArgSerGly 317
 QY 717 CAGTAT----- 722
 Db 318 SerTySerGlyArgSerProSerProTyArgSerProTyArgSerSerSerProPheLeu 337
 QY 723 TCTAAGAAAACCTCAGAGAGCA-----AATCAAGATGTTGCCACGCTCTCACTGG 776
 Db 338 SerIleArgSerLeuSerArgSerProLeuProSerArgIleSerMetLeuSerArgSer 357
 QY 777 GCCTTCCTCTCTCACTCTGATGACACCTCTCTCAATGAAATTCGATTAATCACTCAA 836
 Db 358 ArgSerProAlaTySer----- 363
 QY 837 CGACCAACAGATGCTACACAGAAAG-----AGAAACAGAAATTCGCTCT 881
 Db 364 ArgHisSerSerHisSerIleSerIleSerIleSerIleSerIleSerIleSerIle 881
 QY 882 GTTGGAGGAGCAGAGGTGGAGCTCAGCGCTCTCTCTGTAACACAGAGTTCAAGGAGA 941
 Db 383 rIleSerProValArgLeuProLeuAsnSerSerLeu-----GlyAlaG 398
 QY 942 GCTCGCTGATCCCGAGTCCCATATTACGAGGAGTAGCAGGAAAGTCCCACTTCAGAT 1001
 Db 398 uLeuSerArgIle----- 1001
 QY 1002 GCAAAAGATATTAAAGAACTCCAGGATTCAAAAGATCCATGTTAGATTTAGACC 1061
 Db 408 aAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1061
 QY 1062 AAGAAAGAAAAGATGGCTCAAGCTCCACAGAGATGCACTTACGCGCATTTAAGG 1121
 Db 428 oArgIleGluAsnSer----- 433
 QY 1122 ACACAGTCAAGCAACAAAAGCCCTGCAAGTGCCTCTGTTTGTGATTTCAACAAAAT 1181
 Db 434 ---SerValGluAlaIle-----AspSerGly-----Le 442
 QY 1182 TGAAGTGAAGAGTCTATCATGGAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1241
 Db 442 uGluSerIleSerIleSerIleSerIleSerIleSerIleSerIleSerIleSerIle 1241
 QY 1242 CACAGCTACAGACTCAAAAGGCTGATCAGCAAGCACTAGAGGAGGAGGAGGAGGAGGAG 1301
 Db 460 -----ThrGluLeuValAsnValThrHisLeuAsnThrGluValIleAsnSerSerAs 477
 QY 1302 TGTGGGAGCAATTCAGTCTCAGTAT-----GAAATTCGTCGATCCTCCAGC 1349
 Db 477 pThrGlyValValLeuAspGluAsnSerGluIleValHisLeuValIleAspLeuVal 497
 QY 1350 CTTTGGTCTGACACCAATCAGAGCTGCCACATCTTTCTGTTATACG-----GA 1403
 Db 497 aGlnGlyThrArgAspSerIleValLeuValLeuValLeuValLeuValLeuValLeu 517
 QY 1404 GATGCTACTTGGTCCAGCACTTCTCTGTTGAAACCCAGCTTGCAGACAGTGGACGG 1463
 Db 517 uThrGluThrSerGluValGluThrPro-----ProLeuProThrIleAlaSe 534
 QY 1464 ACAGAGAGTGTCTACCTGACACTTCTTGCTCT-----CCACTGC 1505
 Db 534 rProProProProLeuProToThrThrThrProProProProProProProProPro 554
 QY 1506 TATGGCTCTACCTCTGTCAGAGCTCCACTTCTTATGTCATCAGCATCTCTC 1565
 Db 554 uProProIleProAlaLeuProGlnGlnProProLeuProProSerGlnProAlaPheSe 574
 QY 1566 TCTGACTGATCAAGCACCACAGATACAAATGGCCACTGACGACAAATGCTAGTACCAGG 1625

Db 574 rGlnValProAlaSerSerThrSer----- 582
 QY 1626 GCTCAGCATCCCAACAGATGATTTATTCGCAATCAGCACTG---GCTCTGGGAATTC 1682
 Db 583 ---ThrLeuProProSerThrHisSerIleSerIleSerIleSerIleSerIleSerIle 601
 QY 1683 ACATCCACCT-----GCATCTTCAGATGA 1706
 Db 601 rGlnProProGlnValSerValIleValSerValThrAlaAlaIleProHi 621
 QY 1707 CAGCCGATCAAGTGCAC-----GGTGGCGAGA 1733
 Db 621 eLeuIleThrSerThrLeuProLeuProLeuProLeuProLeuProGlyGlyAspAs 641
 QY 1734 TATG----- 1737
 Db 641 pMetAspSerProIleSerIleThrLeuProSerIleProValIleSerGluIleGlnAr 661
 QY 1738 -GTCAACACCTAGATGAATGATCTGTCGACACTCTGCTCCCATCTGAGGTACCA 1794
 Db 661 gThrArgHisLeu-----LeuThrAspLeuProLeuProGluLeuPro 676
 RESULT 10
 ID PCGV RAT STANDARD; PRT; 2738 AA.
 AC Q9ERB4; O08592; O08564; Q9R1K4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
 DE hyaluronate-binding protein) (GHA) (Fragments).
 GN CP52.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
 RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
 RC STRAIN=Wistar Kyoto;
 RX MEDLINE=9937053; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wight T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [2]
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
 RC STRAIN=Wistar Kyoto;
 RX MEDLINE=98308094; PubMed=9642104;
 RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
 RA Margolis R.K., Margolis R.U.;
 RT "Differential regulation of expression of hyaluronan-binding
 RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
 RT brevican.";
 RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
 RN [3]
 RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
 RC TISSUE=Kidney;
 RX MEDLINE=98094159; PubMed=9434070;
 RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
 RT "Proteoglycan expression in the normal rat kidney.";
 RL Nephron 77:461-470(1997).
 RN [4]
 RP SEQUENCE OF 2535-2738 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Lung;
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
 RT "Molecular cloning and characterization of two developmentally
 RT regulated genes in rat lung.";
 RL Submitted (SEP-2000) to the ENBL/GenBank/DBJ databases.
 CC -I- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds

hyaluronic acid.

-|- SUBUNIT: Interacts with PELNI (By similarity).

-|- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

-|- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;

Comment=Additional isoforms seem to exist;

Name=V0;

Isoid=Q9ERB4-1; Sequence=Displayed;

Name=V3;

Isoid=Q9ERB4-2; Sequence=VSP_003091;

Name=Vint;

Isoid=Q9ERB4-3; Sequence=VSP_003092;

-|- TISSUE SPECIFICITY: In kidney is expressed in the papillary area, but not in glomeruli.

-|- DEVELOPMENTAL STAGE: Disappears after the cartilage development (By similarity).

-|- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

-|- SIMILARITY: Contains 2 link domains.

-|- SIMILARITY: Contains 2 EGF-like domains.

-|- SIMILARITY: Contains 1 C-type lectin family domain.

-|- SIMILARITY: Contains 1 Sushi (SCR) domain.

-|- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/annouce/> or send an email to license@isb-sib.ch).

EMBL; AF062402; AAC40166.1; -

EMBL; U75306; AAB51125.1; -

EMBL; AF084544; AAD48544.1; -

EMBL; AF072892; AAC26116.1; -

EMBL; AY007691; AAG16631.1; -

HSP; P01132; 1BPG.

InterPro; IPR000152; Asx hydroxyl_s.

InterPro; IPR000742; EGF 2.

InterPro; IPR001881; EGF Ca.

InterPro; IPR006209; EGF-like.

InterPro; IPR007110; Ig-like.

InterPro; IPR003599; Ig.

InterPro; IPR001304; Lectin C.

InterPro; IPR000538; Link.

InterPro; IPR000436; Sushi_SCR_CCP.

PRINTS; PR01265; LINKMODULE.

ProDom; PD000918; Link; 2.

SMART; SM00032; CCP; 1.

SMART; SM00034; CLECT; 1.

SMART; SM00179; EGF_CA; 1.

SMART; SM00409; IG; 1.

SMART; SM00445; LINK; 2.

PROSITE; PS00010; ASX HYDROXYL; 1.

PROSITE; PS00615; C TYPE LECTIN_1; 1.

PROSITE; PS50041; C TYPE LECTIN_2; 1.

PROSITE; PS00022; EGF 1; 2.

PROSITE; PS01186; EGF 2; 1.

PROSITE; PS50026; EGF 3; 2.

PROSITE; PS01187; EGF_CA; 1.

PROSITE; PS50835; IG LINK; 1.

PROSITE; PS01241; LINK; 2.

Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi; Signal; Repeat; EGF-like domain; Calcicium; Immunoglobulin domain; Hyaluronic acid; Alternative splicing.

SIGNAL

CHAIN 1 20 POTENTIAL.

FT CHAIN 21 2738

FT NON CONS 348 349

FT DOMAIN 21 146

FT DOMAIN 167 244

FT DOMAIN 265 346

FT DOMAIN <349 695

FT GAG-ALPHA

FT (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).

| | | | |
|----|------|--|------|
| QY | 2012 | AGCTGCTGGTTCATATCATCTACGATCCAACTCTTACAGGATTAAAGCAACTTGAATACTTA | 2071 |
| Db | 2377 | lnArgileuLeuAspTyr-----SerAsnGlnAlaThrValSerThrLeuAsp-----LeuA | 2394 |
| QY | 2072 | ACITTCAGAAACGGG-----AGTGTGATTTGTGAATAGCAAAATCAAGTTTGCTA | 2119 |
| Db | 2394 | snThrGluHiserileProProPheSerileLeuAspAsnSer----- | 2408 |
| QY | 2120 | AGTCTGTGCGGTATAAACCCTCACCAGGCTGTGCACGGGCTCTTGAGGATTTCGTTCTG | 2179 |
| Db | 2409 | -----AsnGluThrAlaPheLeuLeuGlyIleSerGluGluThrValGluG | 2424 |
| QY | 2180 | CTCAGCCCAACAACACTCCATCTCGAAATAGACAGCTACTCTCTCAACATTGAAACACGCTG | 2239 |
| Db | 2424 | lyThrAlaValTyLeu----- | 2430 |
| QY | 2240 | ATCAACGACATCCCTGCAGTTCCTGGCTGCGCGCAATTTGCCCAATGTGTAAGAACG | 2299 |
| Db | 2430 | roGlyProAspLeuCyseLysThrAsnProCyseLysAsnGlyIleThrCysTyrProThrG | 2450 |
| QY | 2300 | AACGGACTCAGGAGCGGAGTCTCGCTGCAAAACCCAGGATAT----- | 2340 |
| Db | 2450 | luThrSer-----TyrValCysThrCyseAlaProGlyTyrSerGlyAspGlnCysGluL | 2468 |
| QY | 2341 | -----GACAGCCAGGAGGAGCCCTGGACGCGTC | 2365 |
| Db | 2468 | euAspPheAspGluCysHisSerAsnProCyseArgAsnGlyAlaThrCyseValAspGlyL | 2488 |
| QY | 2366 | TGGAACCA-----GGCTCTCTGGGCCCTCGCA | 2392 |
| Db | 2488 | euAsnThrPheArgCysLeuCysLeuProSerTyrValGlyAlaLeuCyseGluGlnAspT | 2508 |
| QY | 2393 | CAAGGAATCGAGCTC-----CTCCAGCGAAG-----GGAGCTC | 2428 |
| Db | 2508 | hrGluThrCysAspTyrGlyTrpHisLysPheGlnGlyGlnCysTyrLysTyrPheAlaH | 2528 |
| QY | 2429 | CATGCAGTTCGCACATCACTCTGAAATCAA-----GCATACAAACTA | 2473 |
| Db | 2528 | IsArgArgThrTrpAspAlaGluArgGluCyseArgLeuGlnGlyAlaHisLeuThrS | 2548 |
| QY | 2474 | GTGTAAAGATTCCAAATCAACA----- | 2499 |
| Db | 2548 | erileLeuSerHisGluGluGlnMetPheValAsnArgValGlyHisAspTyrGlnTrpI | 2568 |
| QY | 2500 | -----AATAACAAGGTATCAGTAAA-----AGCAATTCGAATCTACTGACCGTAG | 2545 |
| Db | 2568 | leglyLeuAsnAspLysMetPheGluHisAspPheArgTrpThrAspGlySerAlaLeuG | 2588 |
| QY | 2546 | AATATGAGAATT-----AACCATCAAGAT----- | 2571 |
| Db | 2588 | lnTyrGluAsnTrpArgProAsnGlnProAspSerPhePheSerAlaGlyAspCyseV | 2608 |
| QY | 2572 | -----TGGGAAGGAAATATAAACTGAAA-----ATGTACAAATTCACCTTAG | 2614 |
| Db | 2608 | alValleilelelTrpHisGluAsnGlyGlnTrpAsnAspValProCyseAsnTyHisLeuT | 2628 |
| QY | 2615 | GCTAT-----CTCAAGACGATGATTG-----CCTCTCTCAGGAAATGGA | 2656 |
| Db | 2628 | hrTyrThrCysLysLysGlyThrValAlaCyseGlyGlnProProValValGluAsnAla | 2647 |
| QY | 2657 | GACAGCATATTCATGGGTCAATCAAAATCCAGACATACAGTCAACACTGAGATCAGCAC | 2716 |
| Db | 2648 | -----LysThrPheGlyLysMetLysProArgTyrGluIleAsnSer----- | 2661 |
| QY | 2717 | ACACCATATTCATATAGACAGTCAATGCTGGCAACCACTGAAATCTGAAAAAAA | 2776 |
| Db | 2662 | -----LeuileArgTyrHisCysLysAspGlyPhe-----IleGln | 2673 |
| QY | 2777 | AGACACTTACTTATTATAA----- | 2797 |
| Db | 2674 | ArgHisLeuProThrleArgCysLeuGlyAsnGlyArgTrpAlaMetProLysileThr | 2693 |

```

QY 2798 -----ACCCCAATGCAATCAGCGAACAATAT 2824
      |||:::|
Db 2694 CysMetAenProSerAlaTyGlnArgThrTyr 2704
      |||:::|

RESULT 11
AMNH YEAST
ID AMNH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucosylase SI/S2 precursor (EC 3.2.1.3) (Glucon 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucosylase).
GN STAI OR STAZ OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313266; PubMed=9169870;
RA Churcher C.M., Bowman S., Baddock K., Bankier A., Brown D.,
RA Chillingworth T., Connor K., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA Moulie S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX,"
RT Nature 387:84-87(1997).
RL [2]
EN SEQUENCE FROM N.A.
RP MEDLINE=1794600; PubMed=3106330;
RX Yamashita I., Nakamura M., Fukui S.;
RA "Gene fusion is a possible mechanism underlying the evolution of
RA STAI ";
RT J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF I-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RA "Similar short elements in the 5' regions of the STAZ and SGA genes
RA from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chain
CC with release of beta-D-glucose.
CC -I- SIMILARITY: TO S.POMBE SPBC215.13.
CC -I- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See http://www.isb-sib.ch/annou
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z38061; CAA86176.1; -
CC EMBL; M16164; AAA35014.1; -
CC EMBL; M16165; AAA35015.1; -
CC EMBL; X13857; CAA32069.1; -
CC PIR; S48478; S48478.
CC GerMOnline; 139731; -
CC SGD; S0001458; MUC1
CC GO; GO:0005886; C:plasma membrane; IDA.
CC GO; GO:0030447; P:filamentous growth; IDA.
CC GO; GO:0007125; P:invasive growth; IMP.
CC GO; GO:0007124; P:pseudohyphal growth; IMP.
CC GO:Glossary; Glycosidase; Polysaccharide degradation; Glycoprotein;
CC Signal; Multigene family.
CC 1 21
CC STGNAL POTENTIAL.
CC CHAIN 22 1367
CC GLUCOSYLASE SI/S2.

```

555 roThrProSerSer-----SerThrThrGluSerSerSerThrProValThrSerSerT 573

1601 CTGACCAGACAAATGCTAGTACACGAGGCTCACCATCCCAACAGTGATTTCTGCAATCA 1660

573 hrThrGluSerSerSerAlaPro-----ValProThrProSerSerSerThrThrG 590

1661 GCCAACTGGCTCTGGGAATTTCAATCCACCTGCGATCTTCAGATGACACCGCATCAAGTG 1720

590 luSerSerSerAlaProValProThrProSerSerSerThrThrGluSer----- 606

1721 CAGGTGGCGGAAGATATGGTCAGACACCTAGATGAATGGATCTGTGCACTCTCGCC 1780

607 -----SerSerAlaProAlap 612

1781 CATCTGAGGTACACAGCTCAGCGAATATGTTCTGTCCCGACATATTTCTTGGAGGATA 1840

612 roThrProSerSerSerThrThrGluSerSerSerAlaPro-----ValThrSerSerT 630

1841 CCACCT-----CCTGTCTCAGCTTTACAGTATATCACACT-----AGTT 1879

630 hrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerS 650

1880 CTATGACCAATGCC-----CCAAAGGCGCCGAGAGCTGGTAGTGTCTTCAGTCTGCGTGTG 1936

650 erAla-ProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal--- 668

1937 CTAACATGGCCTTCCCAACGACCTTCCACAGAGCTCTTCGAGTACCGAGCTCTG 1996

669 -----ProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSe 685

1997 AGCAACAATTCACAGCTGCTGGTCCATA-----TCTACGATCCCAATCT 2042

685 rSerThrThrGluSerSerSerAlaProValThrSerSerSerThrThrGluSer 702

RESULT 12

PGCV_HUMAN

ID PGCV_HUMAN STANDARD; PRT; 3396 AA.

AC PI3611; P20754; Q13010; Q13189; Q15123; Q9UNW5;

DT 01-NOV-1990 (Rel. 13, Created)

DT 01-JAN-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Versican core protein precursor (large fibroblast proteoglycan)

DE (Chondroitin sulfate proteoglycan core protein 2) (Glia1

GN hyaluronate-binding protein) (GHP).

GN CSFG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxID=9606;

RN [1]

RX SEQUENCE FROM N.A. (ISOFORM V0).

RX MEDLINE=95105188; PubMed=7528742;

RA Nabe M.F., Zimmermann D.R., Iozzo R.V.;

RT "Characterization of the complete genomic structure of the human

RT versican gene and functional analysis of its promoter.";

RL J. Biol. Chem. 269:32999-33008(1994).

RN [2]

RX SEQUENCE FROM N.A. (ISOFORM V1).

RX TISSUE=Placenta;

RC MEDLINE=90059882; PubMed=2583089;

RX Zimmermann D.R., Ruoslahti E.;

RL "Multiple domains of the large fibroblast proteoglycan, versican.";

RL EMBO J. 8:2975-2981(1989).

RN [3]

RX SEQUENCE FROM N.A. (ISOFORM V2).

RX TISSUE=Glia1 tumor;

RC MEDLINE=95105187; PubMed=7806529;

RX Dours-Zimmermann M.T., Zimmermann D.R.;

RA "A novel glycosaminoglycan attachment domain identified in two

RT alternative splice variants of human versican.";

RL J. Biol. Chem. 269:32992-32998(1994).

RN [4]

RX SEQUENCE OF 2711-3396 FROM N.A.

RC TISSUE=Lung fibroblast;
 RX MEDLINE=88007514; PubMed=2820964;
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 RT lectin-like and growth factor-like sequences.";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Lozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 RT human chromosome 5 (5q12-5q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RX TISSUE=Brain;
 RA MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RX TISSUE=Aortic smooth muscle;
 RA MEDLINE=99327053; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wright T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [8]
 RP PARTIAL SEQUENCE.
 RX TISSUE=Brain;
 RA MEDLINE=89174663; PubMed=2466833;
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial
 RT hyaluronate-binding protein.";
 RL J. Biol. Chem. 264:5981-5987(1989).
 RN [9]
 RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE=96213482; PubMed=8627343;
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -I- FUNCTION: May play a role in intracellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -I- SUBUNIT: Interacts with FBLN1 (By similarity).
 CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=p13611-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=p13611-2; Sequence=VSP_003082, VSP_003083;
 CC Name=V2;
 CC IsoId=p13611-3; Sequence=VSP_003084;
 CC Name=V3;
 CC IsoId=p13611-4; Sequence=VSP_003082, VSP_003085;
 CC Name=Vint;
 CC IsoId=p13611-5; Sequence=VSP_003086;
 CC -I- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; v2 is restricted to normal brain
 CC and gliomas; v3 is found in all these tissues except
 CC medulloblastomas.
 CC -I- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -I- SIMILARITY: Contains 2 link domains.
 CC -I- SIMILARITY: Contains 2 EGF-like domains.

CC -I- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -I- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -I- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U16306; AAA65018.1; -
 CC EMBL: X15988; CAA34128.1; -
 CC EMBL: S52488; AAB24878.1; -
 CC EMBL: U26555; AAA67565.1; -
 CC EMBL: D32039; EAA06801.1; -
 CC EMBL: J02814; AAA36437.1; -
 CC EMBL: AF084545; AAD48545.1; -
 CC PIR: S06014; A60979.
 CC HSSP: P01132; IEGF.
 CC Genew; HGNC:2464; CSPG2.
 CC MIM; 118661; -
 CC GO; GO:0005578; C:extracellular matrix; TAS.
 CC GO; GO:0005540; F:hyaluronic acid binding; TAS.
 CC GO; GO:0008037; P:cell recognition; TAS.
 CC GO; GO:0007275; P:development; TAS.
 CC InterPro: IPR000152; Asx_hydroxyl_S.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR006209; EGF_like.
 CC InterPro: IPR007110; Ig_Like.
 CC InterPro: IPR003599; Ig.
 CC InterPro: IPR001304; Lectin_C.
 CC InterPro: IPR000538; Link.
 CC InterPro: IPR000436; Sushi_SCR_CCP.
 CC Pfam: PF00008; EGF_2.
 CC Pfam: PF00059; lectin_C; 1.
 CC Pfam: PF00084; sushi_1.
 CC Pfam: PF00193; Xlink; 2.
 CC PRINTS; P01265; LINKMODULE.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_CA; 1.
 CC SMART; SM00409; Ig; 1.
 CC SMART; SM00445; Link; 2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 1.
 CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 CC PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
 CC PROSITE; PS00022; EGF_1; 2.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS50026; EGF_3; 2.
 CC PROSITE; PS01187; EGF_CA; 1.
 CC PROSITE; PS00835; IG_Like; 1.
 CC PROSITE; PS01241; LINK; 2.
 CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 CC Hyaluronic acid; Alternative splicing.
 CC SIGNAL 1 20
 CC POTENTIAL.
 CC FT CHAIN 21 3396 VERSICAN CORE PROTEIN.
 CC FT DOMAIN 21 146 IG-LIKE V-TYPE.
 CC FT DOMAIN 167 244 LINK 1.
 CC FT DOMAIN 265 346 LINK 2.
 CC FT DOMAIN 348 1335 GAG-ALPHA
 CC (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
 CC DOMAIN 1336 3089 GAG-BETA.
 CC FT DOMAIN 3089 3125 EGF-LIKE 1.
 CC FT DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 CC FT DOMAIN 3176 3290 C-TYPE LECTIN.
 CC FT DOMAIN 3295 3353 SUSHI.
 CC FT DISULFID 44 130 BY SIMILARITY.
 CC FT DISULFID 172 243 BY SIMILARITY.

```

FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 345 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 3093 3104 BY SIMILARITY.
FT DISULFID 3098 3113 BY SIMILARITY.
FT DISULFID 3131 3124 BY SIMILARITY.
FT DISULFID 3136 3151 BY SIMILARITY.
FT DISULFID 3153 3162 BY SIMILARITY.
FT DISULFID 3169 3180 BY SIMILARITY.
FT DISULFID 3197 3289 BY SIMILARITY.
FT DISULFID 3265 3281 BY SIMILARITY.
FT DISULFID 3296 3339 BY SIMILARITY.
FT DISULFID 3325 3352 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).

Alignment Scores:
Pred. No.: 0.209 Length: 3396
Score: 146.50 Matches: 184
Percent Similarity: 29.66% Conservative: 115
Best Local Similarity: 18.25% Mismatches: 313
Query Match: 2.49% Indels: 396
DB: 1 Gaps: 48

US-10-007-270-1 (1-3330) x PGCV_HUMAN (1-3396)
QY 691 AGCAATCTACATTTCAAAGACTTGGCGAGTATTCTAGAAACCTCAGAAGACCAATT 750
Db 2455 SerSerThrThrPheValSerAspGlySerLeuGluLysHisPro-----GluVal 2471
QY 751 CRAAGTGTGGCAACGCTCACTTGGCGCTTTCCTCTCACT----- 792
Db 2472 ProSerAlaLysAlaValThrAlaAspGlyPheProThrValSerValMetLeuProLeu 2491
QY 793 -----CCTGATGACACCCCTCCCTCAAT----- 813
Db 2492 HisSerGluGlnAsnLysSerSerProAspProThrSerThrLeuSerAsnThrValSer 2511
QY 814 -----GAAATCTCGATAATACA 831
Db 2512 TyrGluArgSerThrAspGlySerPheGlnAspArgPheArgGluPheGluAspSerThr 2531
QY 832 CTCACGACACCAAGATGCTACACAGAA----- 861
Db 2532 LeuLysProAsnArgLysLysProThrGluAsnIleIleAspLeuAspLysGluAsp 2551
QY 862 AGAGAAACAGAAATTCGCTGTGGAGGACGACAGGGTGGAGCTCAGCGTCTCTCGTA 921
Db 2552 LysAspLeuIleLeuThrIleThrGluSerThrIleLeuGluIleLeuProGluLeuThr 2571
QY 922 AACCAAGATTCAAGCGAGCTCGCTGCTCCAGTCCCATATTTACCAGAGGCTAGCA 981
Db 2572 SerAspLysAsnThrIleIleAspIleAspHisThrLysProValTyrGluAspIleLeu 2591
QY 982 GGAAAGTCCCACTTCAGATGCAAAAGATATTAGAAACTTCAGAGTTCAAAAATC 1041
Db 2592 Gly-----MetGlnThrAspIleAspThrGluValProSer----- 2603
QY 1042 CATGTGTAGGATTTAGACAAAGAAAGAAAGAGTGGCTCAAGCTCCACAGAGTCAA 1101
Db 2604 -----GluProHisAspSerAsnAspGluSerAsnAspAspSerThrGln 2618
QY 1102 CTTAGCGGCATCTTTAAGAGACACAGTGCAGAGCAAAAAGCCCTCCAAAGTGAACCTCG 1161
Db 2619 ValGlnGluIleTyrGlu-----AlaAlaValAsnLeu 2629
QY 1162 TCTTTGATTCCAAACAAATTTGAAGTGGAGAGTCTATCATGGAACCACTGGAGGAGC 1221
Db 2630 SerLeu-----ThrGluGluThrPheGluGlySer----- 2639
QY 1222 AAGCAACCAAGAAATCTATCTCAAGCTACAGACCTCAAAAGGCTGATCAGCAAGCACTA 1281

```

```

Db 2640 -----AlaaspValLeuAlaSerTyrThrGlnAlaThrHisAspGluSerMetThrTyr 2657
QY 1282 GAGGAAGAACAAATCTTTGGATGGGGGCAAAATTCAGTTTCACATGATGAAATTCGTGATCA 1341
Db 2658 GluAspArgSerGlnLeuAspHisMetGlyPheHisPheThrThr-----Gly 2673
QY 1342 CTGCCAGCTTTGGTCTCGACACCAATCAGAGCTG-----CCACA 1383
Db 2674 IleProAla-----ProSerThrGluThrGluLeuAspValLeuLeuProThrAlaThr 2691
QY 1384 TCTTTTGTCTTTATAACAGAGGATGCTACTTTGAGTCACAGAACTCTCTCTGTTGACCC 1443
Db 2692 SerLeuProIleProArgLysSerAlaThrValIleProGluIleGluIleLysAla 2711
QY 1444 CAGCTTGACAGACGTGAC----- 1461
Db 2712 GluAlaLysAlaLeuAspMetPheGluSerSerThrLeuSerAspGlyGlnAlaIle 2731
QY 1461 ----- 1461
Db 2732 AlaAspGlnSerGluIleIleProThrLeuGlyGlnPheGluArgThrGlnGluGluTyr 2751
QY 1462 -----GGAGCAGAGCAT 1473
Db 2752 GluAspLysLysHisAlaGlyProSerPheGlnProGluPheSerSerGlyAlaGluGlu 2771
QY 1474 GGTCTA-----CCTGACACTTCTTGTCTCCACCTGCTATGGCTCTTACCTCC 1521
Db 2772 AlaLeuValAspHisThrProTyrLeuSerIleAlaThrHisLeuMetAspGlnSer 2791
QY 1522 CTGTCAAGAGCT-----CCACCTTCTCTT----- 1545
Db 2792 ValThrGluValProAspValMetGluGlySerAsnProTyrTyrThrAspThrThr 2811
QY 1546 ATGGCATCAAGCATCTTCTCTCTGATCAAGGACCAACAGATACATGGCCACT----- 1602
Db 2812 LeuAlaValSerThrPheAlaLysLeuSerSerGluThrProSerSerProLeuThrIle 2831
QY 1603 -----GACCAGACATGCTAGTACGAGG 1626
Db 2832 TyrSerGlySerGluAlaSerGlyHisThrGluIleProGlnProSerAlaLeuProGly 2851
QY 1627 CTCACATCCCCCAAGT----- 1644
Db 2852 IleAspValGlySerSerValMetSerProGlnAspSerPheLysGluIleHisValAsn 2871
QY 1645 -----GATTATTCTGCAATCGCAACTGGCTCTGGGAATTTCCATCCACCTGCA 1695
Db 2872 IleGluAlaThrPheLysProSerSerGluGluTyrLeuHisIleThrGluProSer 2891
QY 1696 TCTTCAGATGACAGCCGATCAAGTCCAGGTGGCGAGAT-----ATGGTCAGACACCTAGAT 1752
Db 2892 LeuSerProAspThrLysLeuGluProSerGluAspAspGlyLysProGluLeuLeuGlu 2911
QY 1753 GAAATGGATCTGTCTGACACTCTCTGCC----- 1779
Db 2912 GluMetGluAlaSerProThrGluLeuIleAlaValGluGlyThrGluIleLeuGlnAsp 2931
QY 1780 -----CCATCT----- 1785
Db 2932 PheGlnAsnLysThrAspGlyGlnValSerGlyGluAlaIleLysMetPheProThrIle 2951
QY 1786 GAGGTACCAGAGCTCAGCGAATATGTTCTTGTCGCCAGATCATTC-----TTGAGAGATACC 1842
Db 2952 LysThrProGluAlaGlyThrValIleThrThrAlaAspGluIleGluLeuGluGlyAla 2971
QY 1843 ACT----- 1845
Db 2972 ThrGlnTrpProHisSerThrSerAlaSerAlaThrTyrGlyValGluAlaGlyValVal 2991
QY 1846 -----CCTGTCTCAGCTTTACAGTATATACCATAGTCTTATGACCATGCC 1893

```

QY 2728 CAATATAGAGAGTCATGCTACTTGCGCAACCAAGTAATAATTCTGAAAAAAGAACACTTACT 2787
Dd 3320 -----LeulleArgTy rHisCysLysAspGlyPhe---lleglnRghIseLupr 3355
QY 2788 TATTATATAA-----ACCCT 2802
Dd 3335 oThrileArgCysleuGlyAsnGlyA-gTrpAlatleProLysileThrCysMetAsnP 3355
QY 2803 AAATGCATCATCAGCAACATAT 2824
Dd 3355 oSerAlaTy rGlnArgThy rTyr 3362
RESULT 13
ID NIBA HUMAN STANDARD; PRT; 928 AA.
AC Q9BZ08; Q8TE15; Q9H593; Q9HY8; Q9HC89;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-OCT-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Niban protein.
GN Clorf24.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20546198; PubMed=110111112;
RA Majima S., Kajino K., Fukuda T., Otsuka F., Hino O.;
RT "A novel gene 'Niban' upregulated in renal carcinogenesis: cloning by
the cDNA-amplified fragment length polymorphism approach.";
RL Jpn. J. Cancer Res. 91:869-874(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21218927; PubMed=11318611;
RA Sood R., Bonner T.I., Malakowaka I., Stephan D.A., Robbins C.M.,
RA Connors T.D., Morgenbesser S.D., Su K., Faruqe M.U., Pinkett H.,
RA Graham C., Baxevasis A.B., Klingler K.W., Landes G.M., Trent J.M.,
RA Carpen J.D.;
RT "Cloning and characterization of 13 novel transcripts and the human
RG8 gene from the lq25 region encompassing the hereditary prostate
cancer (hpc1) locus";
RL Genomics 73:211-222(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Mishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kakui Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Maeuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.J., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Hellon E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.

Db 583 euThrAspLeuLysProThrGlySerAsnGlnAlaSerProAlaArgAlaSerA 603
 QY 1365
 Db 603 laileuProGlyValLeuGlySerGluThrLeuSerAsnGluValPheGlnGluSerG 623
 QY 1366
 Db 623 luGluGluLysGlnProGluValProSerSerLeuAlaLysGlyGluSerLeuSerLeuP 643
 QY 1396
 Db 643 roGlyProSerProProAspGlyThrGluGlnValIleSerArgValAspA 663
 QY 1433 CTGTTGAACCCACCTTGACAGCTGGACGAGCAGCATGGTCTCTACGACACTTCTT 1492
 Db 663 roValValAsnProValAlaThrGluAspThrAla-----GlyLeuProGlyThrCys 681
 QY 1493 GGTCTCCACCTGCTATGGCTCTACTCCCTGTCAGAGCTCCA----- 1536
 Db 681 erSerGluLeuGluPheGlyGlyThrLeuGluAspGluGluProAlaGlnGluLuprog 701
 QY 1537 --CCTTTCTTATGGCATCAAGCATCTCTCTGACTGATCAAGGCACACAGATACAA 1594
 Db 701 luProIleThrAlaSerGlySerLeuLysAlaLeu----- 712
 QY 1595 TGGCCATGACACACAGATGCTAGTACCAGGGTCCACATCCCATCCACAGGATATTCTG 1654
 Db 713 -----ArgLysLeuLeuThrAlaSerValGluValProValAspSerAlaProV 729
 QY 1655 CAATCAGCCAACTGGCTCTGGGAATTTACATCCCATCCGATCTCTGATGACAGCCCGAT 1714
 Db 729 alMetGluGluAspThrAsnGlySerHisValProGlnGluAsnGluGluGluGlu 749
 QY 1715 CAGTGCAGGTGCGAGATATGTCAGACACTAGT-----GAATGGATCTGT 1765
 Db 749 luLysGluProSerGlnAlaAlaAlaIleHisProAspAsnCysGluGluSerGluVal 769
 QY 1766 CTGACACTCTGCCCCATCTCAGGTACCCAG 1797
 Db 769 erGluArgGluAlaGlnProProCysProGlu 779

RESULT 14
 POPC_RALSO
 ID POPC_RALSO STANDARD; PRT: 1024 AA.
 AC Q9RS52;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PopC protein.
 GN POPC OR RSP0875 OR RS01646.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GMI1000;
 RX MEDLINE=20253307; PubMed=10792715;
 RA Quenon M., Timmers A.C.J., Boucher C., Arlat M.;
 RT "Two novel proteins, PopB, which has functional nuclear localization
 RT signals, and PopC, which has a large leucine-rich repeat domain, are
 RT secreted through the Hrp-secretion apparatus of Ralstonia
 RL solanacearum.";
 RL Mol. Microbiol. 36:261-277(2000).
 RP [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=GMI1000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brotter P., Camus J.C., Cattolico L.,

Chandler M., Choisme N., Claudel-Renard C., Cumac S., Demange N.,
 Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,
 Weissenbach J., Boucher C.A.;
 "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 Nature 415:497-502(2002).
 CC -!- FUNCTION: Probably involved in host-pathogen interactions. May
 CC interact with plant target proteins; may modulate a plant signal
 CC transduction pathway. Secreted.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Contains 22 leucine-rich (LRR) repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ245811; CAB57879.1; -;
 DR EMBL: AL646081; CAD18026.1; -;
 DR InterPro: IPR001611; LRR.
 DR Pfam: PF00560; LRR; 11.
 KW Repeat; Leucine-rich repeat; Plasmid; Complete proteome.
 FT REPEAT 192 218 LRR 1.
 FT REPEAT 223 246 LRR 2.
 FT REPEAT 248 270 LRR 3.
 FT REPEAT 271 293 LRR 4.
 FT REPEAT 295 315 LRR 5.
 FT REPEAT 316 338 LRR 6.
 FT REPEAT 340 359 LRR 7.
 FT REPEAT 360 381 LRR 8.
 FT REPEAT 382 403 LRR 9.
 FT REPEAT 404 427 LRR 10.
 FT REPEAT 429 450 LRR 11.
 FT REPEAT 451 475 LRR 12.
 FT REPEAT 477 495 LRR 13.
 FT REPEAT 496 519 LRR 14.
 FT REPEAT 521 543 LRR 15.
 FT REPEAT 544 566 LRR 16.
 FT REPEAT 568 589 LRR 17.
 FT REPEAT 590 612 LRR 18.
 FT REPEAT 614 636 LRR 19.
 FT REPEAT 637 661 LRR 20.
 FT REPEAT 662 687 LRR 21.
 FT REPEAT 824 847 LRR 22.
 FT CONFLICT 198 198 V -> D (IN REF. 1).
 SQ SEQUENCE 1024 AA; 110929 MW; 27AEC679E9A4EB7D CRC64;
 Alignment Scores:
 Pred. No.: 0.169 Length: 1024
 Score: 145.50 Matches: 166
 Percent Similarity: 33.8% Conservative: 106
 Best Local Similarity: 20.67% Mismatches: 257
 Query Match: 2.48% Indels: 275
 DB: 1 Gaps: 33
 US-10-007-270-1 (1-3330) x POPC_RALSO (1-1024)
 QY 675 CCATGTGTCATTTCAACAGCAATCTCATTTCAAGACTTGGGCGAGTATTCTAAGAAACC 734
 Db 6 ProGlnThrValProGlyHisSerProPheThrProLeuPhePhePheThrHisAsn 25
 QY 735 CTCAGAGAGCGAATTCAAGATTTCCACAGTCTCCACAGTCTGGGCGCTTTCTCTCTACTCC 794
 Db 26 LysArgGluAlaArg-HisAspLeuAlaArgLeuSerLeuThrLeuMetProIleLeuPr 45
 QY 795 TGATGACACCTCTCTCAATCAATTCGTGATATACACTCAACGACACAGATGCTGATC 854
 Db 45 OAAG-----LeuPheHisArgThrSerArgThrSerSerAlaAs 58
 QY 855 AACAGAAAGAGAAACA-----GAATTCGCTCT 881

Db 58 pThrGlnArgAspAlaArgThrProProAsnAlaSerProLeuHisGlyGluProGlyArg 78
|||:|||||:|
QY 882 GTTGGAGGACGAGGGTGGAGCTCAGCGTCTCTCTGTAACACCAAGTCTCAAGCAGA 941
|||:|||||:|
Db 78 gThrProArgSerArgGlyGluLeuGlyArgAsnLeuArg-LeuArgSerAsnAlaGlnTr 98
|||:|||||:|
QY 942 GCTCGCTGACTCCAGT----- 958
|||:|||||:|
Db 98 hrSerGlyThrProGlyThrProAlaArgProGlnIleArgAlaSerAlaSerArgThrA 118
-----CCCCATATTACAGAGCTAGCAGCAAGAGTCCCACTTCA 998
QY 118 laProSerThrProGlnHisProGlnGlyThrGlu-GlyThrArgThrValProAsnSer 137
|||:|||||:|
QY 999 -----GATGCAAAAGATATTTAAGAAACTTCCAGGATCAAAAATTCATCGT 1046
|||:|||||:|
Db 138 ProLeuHisAsnAspAla-ArgVal----- 145
QY 1047 GTTAGGATTATAGACCAAGAAAGAAAGATGGCTCAAGCTCCACAGAGATGCAACTTAC 1106
|||:|||||:|
Db 146 -----PheArgGluArgAlaAspHisThrGlyLeuSerAlaTrpArgThrGluMetLe 163
|||:|||||:|
QY 1107 GGCATCTTTAAGAGACACAGTCAGAGCAAGCAAAAGCCCTGCAAGTGCACTCTCTT 1166
|||:|||||:|
Db 163 uThrArgPheIleGluHisSerArgLys----- 172
QY 1167 TGATTCGCAAAATTGAAAGTGAGGAGTCTATCATGGAACCACTGGAGGAGCAAGCA 1226
|||:|||||:|
Db 173 -----HisGlyLeuAlaAsnAspPheGluG 181
QY 1227 ACCGAAATCTATCTCAAGCTACAGACCTCAAAAGGTGTATCAGCAAGCACTAGAGGA 1286
|||:|||||:|
Db 181 nValArgValTr----- 192
QY 1287 AGAACAACTTTGGATGTGGGACAAATTCAGTTCACTGATGAAATGCTGGATCACTGCC 1346
|||:|||||:|
Db 193 -----AspHisLeuLysSerValLeuArg 200
QY 1347 AGCTTTGGTCTGACACCCAA-----TCAGAGTGGCCACATCTTTTGTGTATTAACAGA 1403
|||:|||||:|
Db 200 gMetSerGlyAspSerValGlnLeuLysSerLeuProVal----- 213
QY 1404 GGATGCTACTTTGAGTCAGAACTTCTCTCTGTTGAA-----CCCCA 1445
|||:|||||:|
Db 214 -----ProGluLeuProAspValThrPheGluLeuAlaHisLeuLysAs 228
QY 1446 GCTTGAGACAGTGGAC----- 1461
Db 228 nLeuGluThrValAspCysAspLeuHisAlaLeuProAlaThrLeuGluAsnLeuPheLe 248
QY 1462 -----GGAGCAGAGCAT-----GGTCTACTGTGACACTTCTTG 1493
|||:|||||:|
Db 248 uLeuGluThrLeuSerLeuLysGlyAlaLysAsnPheLysAlaLeuProAspAlaValTr 268
QY 1494 GTCTCCACTGTATGGCTCTACTCTCTGTGCAAGCT-----CCACC 1538
|||:|||||:|
Db 268 pArgLeuProAlaLeuGlnLeuLysLeuSerGluThrGlyLeuLysSerLeuProPr 288
QY 1539 TTTCTTTATGGCATCAAGCATCTTCTCTGACTGATCAAGGCACACAGATACATGGC 1598
|||:|||||:|
Db 288 oValGlyGlySerAlaLeuGlnArgLeuThrIleGluAspSer----- 303
QY 1599 CACTGACACAGACATGCTAGTACCGGGCTCACCTCCACC-----ATGATTTATTC 1652
|||:|||||:|
Db 304 -----ProLeuGluGlnProAlaGlyPheAlaAspLeuAs 316
QY 1653 TGCATCAGCAACTGGCTCTGGGAATTTTCATCCACCTGCACTCTCAGATGACAGCCG 1712
|||:|||||:|
Db 316 pGlnLeuAlaSerLeuSerLeu-----SerAsnThrLysLeuG 329
QY 1713 ATCAAGTGCAGGTGGCGAAGATATGGTCAGACACTAGATGAAATGATCTGTGACAC 1772
|||:|||||:|

Db 329 uLysLeuSerSerGlyIleGlyGlnLeuProAlaLeuLysSerLeuSerLeuGlnAspAs 349
QY 1773 TCCT-----GCCCATCTCAGGTACCGAGACTCAGCAATATGTTTCTGTCCC 1820
|||:|||||:|
Db 349 nProLysLeuGluArgLeuProLysSerLeuGlyGlnValGluGluLeuThrLeuIleG 369
|||:|||||:|
QY 1821 AGAT-----CATTTCTTGCGATACACCTCTCTCAGCTTTCAGGTATATACCCAC 1874
|||:|||||:|
Db 369 yGlyArgIleHisAlaLeuProSerAlaSerGlyMetSerSerLeuGlnLysLeuThrVa 389
|||:|||||:|
QY 1875 TAGTTCTATGACCATTTGCCCAAGGCGAGAGCTGGTAGTGTCTTTCAGTCTGCGTGT 1934
|||:|||||:|
Db 389 lAspAsnSerSerLeuAlaLysLeuProAlaAspPheGlyAlaLeuGlyAsnLeu----- 407
QY 1935 TGCTAACATGGCTTCTCCAAC-----GACCTGTTCAACAA 1970
|||:|||||:|
Db 408 -AlaHisValSerLeuSerAsnThrLysLeuArgAspLeuProAlaSerIleGlyAsnLe 427
|||:|||||:|
QY 1971 GAGCTCTCTGAGTACCGAGCTCTGAGCAACAATTCACACAGCTGCTGGTTCATATCT 2030
|||:|||||:|
Db 427 uPheThrLeuLysThrLeuSerLeuGlnAspAsnProLysLeuGlySerLeuProAlaSe 447
QY 2031 ACATCCATCTTACAGATTTTAAGCAACTTGAATTACTTAATTCAGAAACGGAGTGT 2090
|||:|||||:|
Db 447 rPheGlyGlnLeuSerGlyLeuGlnGluLeuThrLeu-----AsnGlyAsnAr 463
QY 2091 GATTGTGAATAGCAAAATGAATGTTGCTAAGTCTGTGCGGTATTAACCTCAC----- 2142
|||:|||||:|
Db 463 gIleHisGluLeuProSerMetGlyAlaSerSerLeuGlnThrLeuThrValAspAs 483
QY 2143 -AAGGCTGTGACGGGTCTTGGAGATTTTCTGCTGCGAGCCCAACAACCTCCATCT 2201
|||:|||||:|
Db 483 pThrAlaLeuAlaGlyLeuProAlaAspPheGlyAla-LeuArgAsn----- 498
QY 2202 GGAATACAGACGACTACTCTCTCACTTTGAACCCAGCTGATCAGCAGATCCCTGCAAGT 2261
|||:|||||:|
Db 499 -----LeuAlaHisLeuSerSerAsnThrGlnLeuArgGluLeuProAlaAsn 516
QY 2262 CTGCGCTGCGGCAATTTGCCAATGTGTAAAGACGACGAGTCTGAGGACGGAGCTG 2321
|||:|||||:|
Db 516 hr-----GlyAsnLeu-----HisAlaLeuLysThrLeuSerL 527
QY 2322 TCGCTGCAACACAGATATGAC-----AGCCAGGCGAGCTGCGAGCTCTGG 2368
|||:|||||:|
Db 527 euGlnGlyAsnGlnLeuAlaThrLeuProSerSerLeuGlyTyThrLeuSerGlyLeuG 547
QY 2369 AA-----CCAGGCTCTGTGCGCTGSCA 2392
|||:|||||:|
Db 547 lGluLeuThrLeuLysAsnSerSerValSerGluLeuProProMet---GlyProGlyS 566
QY 2393 CAAGGAATGCGAGTCTCTCCAGGGAAGGGAGCTCCA----- 2430
|||:|||||:|
Db 566 er---AlaLeuLysThrLeuThrValGluAsnSerProLeuThrSerIleProAlaAspI 585
QY 2431 -----TCAGGTGGCAGATCATCTCTGAAATCAAGCATACAAAATAGTGTGA 2479
|||:|||||:|
Db 585 leGlyIleGlnCysGluArg-LeuThrGlnLeuSerLeuSerAsnThrGlnLeuArgAla 604
QY 2480 AAAAGTTCCAAATCAACAAATAACAGGTATCATGTAAGAAATTCGATTAATCTGA 2539
|||:|||||:|
Db 605 LeuProSerSerIleGlyLysLeuSer----- 613
QY 2540 CCGTAGAATATCAAGATTTTAACCATCAAGATTGGAGGAATTAATAATGTAATGT 2599
|||:|||||:|
Db 614 -----AsnLeuLysGlyLeuThrLeuLysAsnAsnAlaArgLeuGlu----- 627
QY 2600 ACAATTAATCATAGGTATCTCAGAGAGATGATTTGCCTTCTCAGGAAATATGGAGAC 2659
|||:|||||:|
Db 628 -----LeuLeuSerGluSerGlyVal 634
QY 2660 AGGCATATTATGGGTATCAAAATCCAGACATACAGTCAACCTGAGAAATCAGAC 2714
|||:|||||:|
Db 635 ArgLysLeuGluSerValArgLysIle-AspLeuSerGlyCysValArgLeuThr 652

RESULT 15
MUC1_MOUSE
ID MUC1_MOUSE STANDARD; PRT: 630 AA.
AC Q02496;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor (Polymorphic epithelial mucin) (PEMT) (Episialin).
GN MUC1 OR MUC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1] _SEQUENCE FROM N.A.
RP MEDLINE=91332029; PubMed=1714452;
RX Spicer A.P., Parry G., Patton S., Gendler S.J.;
RT "Molecular cloning and analysis of the mouse homologue of the tumor-associated mucin, MUC1, reveals conservation of potential O-glycosylation sites, transmembrane, and cytoplasmic domains and a loss of minisatellite-like polymorphism.";
RT J. Biol. Chem. 266:15099-15109(1991).
RN [2] _SEQUENCE FROM N.A.
RP MEDLINE=92068178; PubMed=1958179;
RX Vos H.L., Devries Y., Hilkens J.;
RT "The mouse episialin (Muc1) gene and its promoter: rapid evolution of the repetitive domain in the protein.";
RT Biochem. Biophys. Res. Commun. 181:121-130(1991).
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN CYTOSKELETON.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXCLUSIVELY LOCATED IN THE APICAL DOMAIN OF THE PLASMA MEMBRANE OF HIGHLY POLARIZED EPITHELIAL CELLS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF EPITHELIAL TISSUES. ABERRANTLY EXPRESSED IN EPITHELIAL CARCINOMAS.
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84683; AAA39755.1; -;
DR EMBL; U16175; AAA98538.1; -;
DR EMBL; M65132; AAA39755.1; -;
DR EMBL; M64928; AAA39755.1; JOINED.
DR EMBL; M77226; AAA39754.1; -;
DR PIR; A39344; A39344.
DR FIR; I52257; I52257.
DR MGD; MGI:97231; Muc1.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; P550024; SEA; 1.
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane; Repeat.
FT SIGNAL 1 20
FT CHAIN 21 630
FT DOMAIN 21 535
FT TRANSMEM 536 556
FT DOMAIN 557 630
FT DOMAIN 42 366
FT REPEAT 42 61
FT REPEAT 62 81
FT REPEAT 82 101
FT REPEAT 102 122
FT REPEAT 123 143

144 REPEAT 164
145 REPEAT 164
146 REPEAT 164
147 REPEAT 164
148 REPEAT 164
149 REPEAT 164
150 REPEAT 164
151 REPEAT 164
152 REPEAT 164
153 REPEAT 164
154 REPEAT 164
155 REPEAT 164
156 REPEAT 164
157 REPEAT 164
158 REPEAT 164
159 REPEAT 164
160 REPEAT 164
161 REPEAT 164
162 REPEAT 164
163 REPEAT 164
164 REPEAT 164
165 REPEAT 164
166 REPEAT 164
167 REPEAT 164
168 REPEAT 164
169 REPEAT 164
170 REPEAT 164
171 REPEAT 164
172 REPEAT 164
173 REPEAT 164
174 REPEAT 164
175 REPEAT 164
176 REPEAT 164
177 REPEAT 164
178 REPEAT 164
179 REPEAT 164
180 REPEAT 164
181 REPEAT 164
182 REPEAT 164
183 REPEAT 164
184 REPEAT 164
185 REPEAT 164
186 REPEAT 164
187 REPEAT 164
188 REPEAT 164
189 REPEAT 164
190 REPEAT 164
191 REPEAT 164
192 REPEAT 164
193 REPEAT 164
194 REPEAT 164
195 REPEAT 164
196 REPEAT 164
197 REPEAT 164
198 REPEAT 164
199 REPEAT 164
200 REPEAT 164
201 REPEAT 164
202 REPEAT 164
203 REPEAT 164
204 REPEAT 164
205 REPEAT 164
206 REPEAT 164
207 REPEAT 164
208 REPEAT 164
209 REPEAT 164
210 REPEAT 164
211 REPEAT 164
212 REPEAT 164
213 REPEAT 164
214 REPEAT 164
215 REPEAT 164
216 REPEAT 164
217 REPEAT 164
218 REPEAT 164
219 REPEAT 164
220 REPEAT 164
221 REPEAT 164
222 REPEAT 164
223 REPEAT 164
224 REPEAT 164
225 REPEAT 164
226 REPEAT 164
227 REPEAT 164
228 REPEAT 164
229 REPEAT 164
230 REPEAT 164
231 REPEAT 164
232 REPEAT 164
233 REPEAT 164
234 REPEAT 164
235 REPEAT 164
236 REPEAT 164
237 REPEAT 164
238 REPEAT 164
239 REPEAT 164
240 REPEAT 164
241 REPEAT 164
242 REPEAT 164
243 REPEAT 164
244 REPEAT 164
245 REPEAT 164
246 REPEAT 164
247 REPEAT 164
248 REPEAT 164
249 REPEAT 164
250 REPEAT 164
251 REPEAT 164
252 REPEAT 164
253 REPEAT 164
254 REPEAT 164
255 REPEAT 164
256 REPEAT 164
257 REPEAT 164
258 REPEAT 164
259 REPEAT 164
260 REPEAT 164
261 REPEAT 164
262 REPEAT 164
263 REPEAT 164
264 REPEAT 164
265 REPEAT 164
266 REPEAT 164
267 REPEAT 164
268 REPEAT 164
269 REPEAT 164
270 REPEAT 164
271 REPEAT 164
272 REPEAT 164
273 REPEAT 164
274 REPEAT 164
275 REPEAT 164
276 REPEAT 164
277 REPEAT 164
278 REPEAT 164
279 REPEAT 164
280 REPEAT 164
281 REPEAT 164
282 REPEAT 164
283 REPEAT 164
284 REPEAT 164
285 REPEAT 164
286 REPEAT 164
287 REPEAT 164
288 REPEAT 164
289 REPEAT 164
290 REPEAT 164
291 REPEAT 164
292 REPEAT 164
293 REPEAT 164
294 REPEAT 164
295 REPEAT 164
296 REPEAT 164
297 REPEAT 164
298 REPEAT 164
299 REPEAT 164
300 REPEAT 164
301 REPEAT 164
302 REPEAT 164
303 REPEAT 164
304 REPEAT 164
305 REPEAT 164
306 REPEAT 164
307 REPEAT 164
308 REPEAT 164
309 REPEAT 164
310 REPEAT 164
311 REPEAT 164
312 REPEAT 164
313 REPEAT 164
314 REPEAT 164
315 REPEAT 164
316 REPEAT 164
317 REPEAT 164
318 REPEAT 164
319 REPEAT 164
320 REPEAT 164
321 REPEAT 164
322 REPEAT 164
323 REPEAT 164
324 REPEAT 164
325 REPEAT 164
326 REPEAT 164
327 REPEAT 164
328 REPEAT 164
329 REPEAT 164
330 REPEAT 164
331 REPEAT 164
332 REPEAT 164
333 REPEAT 164
334 REPEAT 164
335 REPEAT 164
336 REPEAT 164
337 REPEAT 164
338 REPEAT 164
339 REPEAT 164
340 REPEAT 164
341 REPEAT 164
342 REPEAT 164
343 REPEAT 164
344 REPEAT 164
345 REPEAT 164
346 REPEAT 164
347 REPEAT 164
348 REPEAT 164
349 REPEAT 164
350 REPEAT 164
351 REPEAT 164
352 REPEAT 164
353 REPEAT 164
354 REPEAT 164
355 REPEAT 164
356 REPEAT 164
357 REPEAT 164
358 REPEAT 164
359 REPEAT 164
360 REPEAT 164
361 REPEAT 164
362 REPEAT 164
363 REPEAT 164
364 REPEAT 164
365 REPEAT 164
366 REPEAT 164
367 REPEAT 164
368 REPEAT 164
369 REPEAT 164
370 REPEAT 164
371 REPEAT 164
372 REPEAT 164
373 REPEAT 164
374 REPEAT 164
375 REPEAT 164
376 REPEAT 164
377 REPEAT 164
378 REPEAT 164
379 REPEAT 164
380 REPEAT 164
381 REPEAT 164
382 REPEAT 164
383 REPEAT 164
384 REPEAT 164
385 REPEAT 164
386 REPEAT 164
387 REPEAT 164
388 REPEAT 164
389 REPEAT 164
390 REPEAT 164
391 REPEAT 164
392 REPEAT 164
393 REPEAT 164
394 REPEAT 164
395 REPEAT 164
396 REPEAT 164
397 REPEAT 164
398 REPEAT 164
399 REPEAT 164
400 REPEAT 164
401 REPEAT 164
402 REPEAT 164
403 REPEAT 164
404 REPEAT 164
405 REPEAT 164
406 REPEAT 164
407 REPEAT 164
408 REPEAT 164
409 REPEAT 164
410 REPEAT 164
411 REPEAT 164
412 REPEAT 164
413 REPEAT 164
414 REPEAT 164
415 REPEAT 164
416 REPEAT 164
417 REPEAT 164
418 REPEAT 164
419 REPEAT 164
420 REPEAT 164
421 REPEAT 164
422 REPEAT 164
423 REPEAT 164
424 REPEAT 164
425 REPEAT 164
426 REPEAT 164
427 REPEAT 164
428 REPEAT 164
429 REPEAT 164
430 REPEAT 164
431 REPEAT 164
432 REPEAT 164
433 REPEAT 164
434 REPEAT 164
435 REPEAT 164
436 REPEAT 164
437 REPEAT 164
438 REPEAT 164
439 REPEAT 164
440 REPEAT 164
441 REPEAT 164
442 REPEAT 164
443 REPEAT 164
444 REPEAT 164
445 REPEAT 164
446 REPEAT 164
447 REPEAT 164
448 REPEAT 164
449 REPEAT 164
450 REPEAT 164
451 REPEAT 164
452 REPEAT 164
453 REPEAT 164
454 REPEAT 164
455 REPEAT 164
456 REPEAT 164
457 REPEAT 164
458 REPEAT 164
459 REPEAT 164
460 REPEAT 164
461 REPEAT 164
462 REPEAT 164
463 REPEAT 164
464 REPEAT 164
465 REPEAT 164
466 REPEAT 164
467 REPEAT 164
468 REPEAT 164
469 REPEAT 164
470 REPEAT 164
471 REPEAT 164
472 REPEAT 164
473 REPEAT 164
474 REPEAT 164
475 REPEAT 164
476 REPEAT 164
477 REPEAT 164
478 REPEAT 164
479 REPEAT 164
480 REPEAT 164
481 REPEAT 164
482 REPEAT 164
483 REPEAT 164
484 REPEAT 164
485 REPEAT 164
486 REPEAT 164
487 REPEAT 164
488 REPEAT 164
489 REPEAT 164
490 REPEAT 164
491 REPEAT 164
492 REPEAT 164
493 REPEAT 164
494 REPEAT 164
495 REPEAT 164
496 REPEAT 164
497 REPEAT 164
498 REPEAT 164
499 REPEAT 164
500 REPEAT 164
501 REPEAT 164
502 REPEAT 164
503 REPEAT 164
504 REPEAT 164
505 REPEAT 164
506 REPEAT 164
507 REPEAT 164
508 REPEAT 164
509 REPEAT 164
510 REPEAT 164
511 REPEAT 164
512 REPEAT 164
513 REPEAT 164
514 REPEAT 164
515 REPEAT 164
516 REPEAT 164
517 REPEAT 164
518 REPEAT 164
519 REPEAT 164
520 REPEAT 164
521 REPEAT 164
522 REPEAT 164
523 REPEAT 164
524 REPEAT 164
525 REPEAT 164
526 REPEAT 164
527 REPEAT 164
528 REPEAT 164
529 REPEAT 164
530 REPEAT 164
531 REPEAT 164
532 REPEAT 164
533 REPEAT 164
534 REPEAT 164
535 REPEAT 164
536 REPEAT 164
537 REPEAT 164
538 REPEAT 164
539 REPEAT 164
540 REPEAT 164
541 REPEAT 164
542 REPEAT 164
543 REPEAT 164
544 REPEAT 164
545 REPEAT 164
546 REPEAT 164
547 REPEAT 164
548 REPEAT 164
549 REPEAT 164
550 REPEAT 164
551 REPEAT 164
552 REPEAT 164
553 REPEAT 164
554 REPEAT 164
555 REPEAT 164
556 REPEAT 164
557 REPEAT 164
558 REPEAT 164
559 REPEAT 164
560 REPEAT 164
561 REPEAT 164
562 REPEAT 164
563 REPEAT 164
564 REPEAT 164
565 REPEAT 164
566 REPEAT 164
567 REPEAT 164
568 REPEAT 164
569 REPEAT 164
570 REPEAT 164
571 REPEAT 164
572 REPEAT 164
573 REPEAT 164
574 REPEAT 164
575 REPEAT 164
576 REPEAT 164
577 REPEAT 164
578 REPEAT 164
579 REPEAT 164
580 REPEAT 164
581 REPEAT 164
582 REPEAT 164
583 REPEAT 164
584 REPEAT 164
585 REPEAT 164
586 REPEAT 164
587 REPEAT 164
588 REPEAT 164
589 REPEAT 164
590 REPEAT 164
591 REPEAT 164
592 REPEAT 164
593 REPEAT 164
594 REPEAT 164
595 REPEAT 164
596 REPEAT 164
597 REPEAT 164
598 REPEAT 164
599 REPEAT 164
600 REPEAT 164
601 REPEAT 164
602 REPEAT 164
603 REPEAT 164
604 REPEAT 164
605 REPEAT 164
606 REPEAT 164
607 REPEAT 164
608 REPEAT 164
609 REPEAT 164
610 REPEAT 164
611 REPEAT 164
612 REPEAT 164
613 REPEAT 164
614 REPEAT 164
615 REPEAT 164
616 REPEAT 164
617 REPEAT 164
618 REPEAT 164
619 REPEAT 164
620 REPEAT 164
621 REPEAT 164
622 REPEAT 164
623 REPEAT 164
624 REPEAT 164
625 REPEAT 164
626 REPEAT 164
627 REPEAT 164
628 REPEAT 164
629 REPEAT 164
630 REPEAT 164
631 REPEAT 164
632 REPEAT 164
633 REPEAT 164
634 REPEAT 164
635 REPEAT 164
636 REPEAT 164
637 REPEAT 164
638 REPEAT 164
639 REPEAT 164
640 REPEAT 164
641 REPEAT 164
642 REPEAT 164
643 REPEAT 164
644 REPEAT 164
645 REPEAT 164
646 REPEAT 164
647 REPEAT 164
648 REPEAT 164
649 REPEAT 164
650 REPEAT 164
651 REPEAT 164
652 REPEAT 164
653 REPEAT 164
654 REPEAT 164
655 REPEAT 164
656 REPEAT 164
657 REPEAT 164
658 REPEAT 164
659 REPEAT 164
660 REPEAT 164
661 REPEAT 164
662 REPEAT 164
663 REPEAT 164
664 REPEAT 164
665 REPEAT 164
666 REPEAT 164
667 REPEAT 164
668 REPEAT 164
669 REPEAT 164
670 REPEAT 164
671 REPEAT 164
672 REPEAT 164
673 REPEAT 164
674 REPEAT 164
675 REPEAT 164
676 REPEAT 164
677 REPEAT 164
678 REPEAT 164
679 REPEAT 164
680 REPEAT 164
681 REPEAT 164
682 REPEAT 164
683 REPEAT 164
684 REPEAT 164
685 REPEAT 164
686 REPEAT 164
687 REPEAT 164
688 REPEAT 164
689 REPEAT 164
690 REPEAT 164
691 REPEAT 164
692 REPEAT 164
693 REPEAT 164
694 REPEAT 164
695 REPEAT 164
696 REPEAT 164
697 REPEAT 164
698 REPEAT 164
699 REPEAT 164
700 REPEAT 164
701 REPEAT 164
702 REPEAT 164
703 REPEAT 164
704 REPEAT 164
705 REPEAT 164
706 REPEAT 164
707 REPEAT 164
708 REPEAT 164
709 REPEAT 164
710 REPEAT 164
711 REPEAT 164
712 REPEAT 164
713 REPEAT 164
714 REPEAT 164
715 REPEAT 164
716 REPEAT 164
717 REPEAT 164
718 REPEAT 164
719 REPEAT 164
720 REPEAT 164
721 REPEAT 164
722 REPEAT 164
723 REPEAT 164
724 REPEAT 164
725 REPEAT 164
726 REPEAT 164
727 REPEAT 164
728 REPEAT 164
729 REPEAT 164
730 REPEAT 164
731 REPEAT 164
732 REPEAT 164
733 REPEAT 164
734 REPEAT 164
735 REPEAT 164
736 REPEAT 164
737 REPEAT 164
738 REPEAT 164
739 REPEAT 164
740 REPEAT 164
741 REPEAT 164
742 REPEAT 164
743 REPEAT 164
744 REPEAT 164
745 REPEAT 164
746 REPEAT 164
747 REPEAT 164
748 REPEAT 164
749 REPEAT 164
750 REPEAT 164
751 REPEAT 164
752 REPEAT 164
753 REPEAT 164
754 REPEAT 164
755 REPEAT 164
756 REPEAT 164
757 REPEAT 164
758 REPEAT 164
759 REPEAT 164
760 REPEAT 164
761 REPEAT 164
762 REPEAT 164
763 REPEAT 164
764 REPEAT 164
765 REPEAT 164
766 REPEAT 164
767 REPEAT 164
768 REPEAT 164
769 REPEAT 164
770 REPEAT 164
771 REPEAT 164
772 REPEAT 164
773 REPEAT 164
774 REPEAT 164
775 REPEAT 164
776 REPEAT 164
777 REPEAT 164
778 REPEAT 164
779 REPEAT 164
780 REPEAT 164
781 REPEAT 164
782 REPEAT 164
783 REPEAT 164
784 REPEAT 164
785 REPEAT 164
786 REPEAT 164
787 REPEAT 164
788 REPEAT 164
789 REPEAT 164
790 REPEAT 164
791 REPEAT 164
792 REPEAT 164
793 REPEAT 164
794 REPEAT 164
795 REPEAT 164
796 REPEAT 164
797 REPEAT 164
798 REPEAT 164
799 REPEAT 164
800 REPEAT 164
801 REPEAT 164
802 REPEAT 164
803 REPEAT 164
804 REPEAT 164
805 REPEAT 164
806 REPEAT 164
807 REPEAT 164
808 REPEAT 164
809 REPEAT 164
810 REPEAT 164
811 REPEAT 164
812 REPEAT 164
813 REPEAT 164
814 REPEAT 164
815 REPEAT 164
816 REPEAT 164
817 REPEAT 164
818 REPEAT 164
819 REPEAT 164
820 REPEAT 164
821 REPEAT 164
822 REPEAT 164
823 REPEAT 164
824 REPEAT 164
825 REPEAT 164
826 REPEAT 164
827 REPEAT 164
828 REPEAT 164
829 REPEAT 164
830 REPEAT 164
831 REPEAT 164
832 REPEAT 164
833 REPEAT 164
834 REPEAT 164
835 REPEAT 164
836 REPEAT 164
837 REPEAT 164
838 REPEAT 164
839 REPEAT 164
840 REPEAT 164
841 REPEAT 164
842 REPEAT 164
843 REPEAT 164
844 REPEAT 164
845 REPEAT 164
846 REPEAT 164
847 REPEAT 164
848 REPEAT 164
849 REPEAT 164
850 REPEAT 164
851 REPEAT 164
852 REPEAT 164
853 REPEAT 164
854 REPEAT 164
855 REPEAT 164
856 REPEAT 164
857 REPEAT 164
858 REPEAT 164
859 REPEAT 164
860 REPEAT 164
861 REPEAT 164
862 REPEAT 164
863 REPEAT 164
864 REPEAT 164
865 REPEAT 164
866 REPEAT 164
867 REPEAT 164
868 REPEAT 164
869 REPEAT 164
870 REPEAT 164
871 REPEAT 164
872 REPEAT 164
873 REPEAT 164
874 REPEAT 164
875 REPEAT 164
876 REPEAT 164
877 REPEAT 164
878 REPEAT 164
879 REPEAT 164
880 REPEAT 164
881 REPEAT 164
882 REPEAT 164
883 REPEAT 164
884 REPEAT 164
885 REPEAT 164
886 REPEAT 164
887 REPEAT 164
888 REPEAT 164
889 REPEAT 164
890 REPEAT 164
891 REPEAT 164
892 REPEAT 164
893 REPEAT 164
894 REPEAT 164
895 REPEAT 164
896 REPEAT 164
897 REPEAT 164
898 REPEAT 164
899 REPEAT 164
900 REPEAT 164
901 REPEAT 164
902 REPEAT 164
903 REPEAT 164
904 REPEAT 164
905 REPEAT 164
906 REPEAT 164
907 REPEAT 164
908 REPEAT 164
909 REPEAT 164
910 REPEAT 164
911 REPEAT 164
912 REPEAT 164
913 REPEAT 164
914 REPEAT 164
915 REPEAT 164
916 REPEAT 164
917 REPEAT 164
918 REPEAT 164
919 REPEAT 164
920 REPEAT 164
921 REPEAT 164
922 REPEAT 164
923 REPEAT 164
924 REPEAT 164
925 REPEAT 164
926 REPEAT 164
927 REPEAT 164
928 REPEAT 164
929 REPEAT 164
930 REPEAT 164
931 REPEAT 164
932 REPEAT 164
933 REPEAT 164
934 REPEAT 164
935 REPEAT 164
936 REPEAT 164
937 REPEAT 164
938 REPEAT 164
939 REPEAT 164
940 REPEAT 164
941 REPEAT 164
942 REPEAT 164
943 REPEAT 164
944 REPEAT 164
945 REPEAT 164
946 REPEAT 164
947 REPEAT 164
948 REPEAT 164
949 REPEAT 164
950 REPEAT 164
951 REPEAT 164
952 REPEAT 164
953 REPEAT 164
954 REPEAT 164
955 REPEAT 164
956 REPEAT 164
957 REPEAT 164
958 REPEAT 164
959 REPEAT 164
960 REPEAT 164
961 REPEAT 164
962 REPEAT 164
963 REPEAT 164
964 REPEAT 164
965 REPEAT 164
966 REPEAT 164
967 REPEAT 164
968 REPEAT 164
969 REPEAT 164
970 REPEAT 164
971 REPEAT 164
972 REPEAT 164
973 REPEAT 164
974 REPEAT 164
975 REPEAT 164
976 REPEAT 164
977 REPEAT 164
978 REPEAT 164
979 REPEAT 164
980 REPEAT 164
981 REPEAT 164
982 REPEAT 164
983 REPEAT 164
984 REPEAT 164
985 REPEAT 164
986 REPEAT 164
987 REPEAT 164
988 REPEAT 164
989 REPEAT 164
990 REPEAT 164
991 REPEAT 164
992 REPEAT 164
993 REPEAT 164
994 REPEAT 164
995 REPEAT 164
996 REPEAT 164
997 REPEAT 164
998 REPEAT 164
999 REPEAT 164
1000 REPEAT 164
1001 REPEAT 164
1002 REPEAT 164
1003 REPEAT 164
1004 REPEAT 164
1005 REPEAT 164
1006 REPEAT 164
1007 REPEAT 164
1008 REPEAT 164
1009 REPEAT 164
1010 REPEAT 164
1011 REPEAT 164
1012 REPEAT 164
1013 REPEAT 164
1014 REPEAT 164
1015 REPEAT 164
1016 REPEAT 164
1017 REPEAT 164
1018 REPEAT 164
1019 REPEAT 164
1020 REPEAT 164
1021 REPEAT 164
1022 REPEAT 164
1023 REPEAT 164
1024 REPEAT 164
1025 REPEAT 164
1026 REPEAT 164
1027 REPEAT 164
1028 REPEAT 164
1029 REPEAT 164
1030 REPEAT 164
1031 REPEAT 164
1032 REPEAT 164
1033 REPEAT 164
1034 REPEAT 164
1035 REPEAT 164
1036 REPEAT 164
1037 REPEAT 164
1038 REPEAT 164
1039 REPEAT 164
1040 REPEAT 164
1041 REPEAT 164
1042 REPEAT 164
1043 REPEAT 164
1044 REPEAT 164
1045 REPEAT 164
1046 REPEAT 164
1047 REPEAT 164
1048 REPEAT 164
1049 REPEAT 164
1050 REPEAT 164
1051 REPEAT 164
1052 REPEAT 164
1053 REPEAT 164
1054 REPEAT 164
1055 REPEAT 164
1056 REPEAT 164
1057 REPEAT 164
1058 REPEAT 164
1059 REPEAT 164
1060 REPEAT 164
1061 REPEAT 164
1062 REPEAT 164
1063 REPEAT 164
1064 REPEAT 164
1065 REPEAT 164
1066 REPEAT 164
1067 REPEAT 164
1068 REPEAT 164
1069 REPEAT 164
1070 REPEAT 164
1071 REPEAT 164
1072 REPEAT 164
1073 REPEAT 164
1074 REPEAT 164
1075 REPEAT 164
1076 REPEAT 164
1077 REPEAT 164
1078 REPEAT 164
1079 REPEAT 164
1080 REPEAT 164
1081 REPEAT 164
1082 REPEAT 164
1083 REPEAT 164
1084 REPEAT 164
1085 REPEAT 164
1086 REPEAT 164
1087 REPEAT 164
1088 REPEAT 164
1089 REPEAT 164
1090 REPEAT 164
1091 REPEAT 164
1092 REPEAT 164
1093 REPEAT 164
1094 REPEAT 164
1095 REPEAT 164
1096 REPEAT 164
1097 REPEAT 164
1098 REPEAT 164
1099 REPEAT 164
1100 REPEAT 164
1101 REPEAT 164
1102 REPEAT 164
1103 REPEAT 164
1104 REPEAT 164
1105 REPEAT 164
1106 REPEAT 164
1107 REPEAT 164
1108 REPEAT 164
1109 REPEAT 164
1110 REPEAT 164
1111 REPEAT 164
1112 REPEAT 164
1113 REPEAT 164
1114 REPEAT 164
1115 REPEAT 164
1116 REPEAT 164
1117 REPEAT 164
1118 REPEAT 164
1119 REPEAT 164
1120 REPEAT 164
1121 REPEAT 164
1122 REPEAT 164
1123 REPEAT 164
1124 REPEAT 164
1125 REPEAT 164
1126 REPEAT 164
1127 REPEAT 164
1128 REPEAT 164
112

```
QY 1429 CCTCTGTTGAACCCAGCTTGAGACAGTGGACGAGCAGCATGGTCTACCTGACACT 1488
Db 158 ProAlaThrSerPro-----ProGlyAsp 165
QY 1489 TCTGTGCTCCACCTGCTATGCGCTCTACCTCC-----CTGTCAAGAGCTCCACCTTTC 1542
Db 166 SerThrSerProAspHisSerThrSerProAlaThrArgAlaProGluAsp 185
QY 1543 TTTATGGCATCAAGCATCTCTCTGACTGAT-----CAA 1578
Db 186 SerThrSerThrAlaValLeuSerGlyThrSerSerProAlaThrThrAlaProValAsp 205
QY 1579 GGCACACACAGATCAATGCGCCACTGACACAGACATCTAGTACCAGGCGCTCCACCTCCCC 1638
Db 206 SerThrSerProValAlaHisAspThrSer---SerProAlaThrSerLeu--- 223
QY 1639 ACCAGTGATTTCTGCAATCAGCCCACTGGCTCTGGGAATTTACATCCACCTGCTACT 1698
Db 224 ---SerGluAspSerAlaSerSerProValAlaHisGlyThrSerSerProAlaThr 242
QY 1699 TCAGAT-----GACAGCCGATCAAGTCCA----- 1722
Db 243 SerProLeuArgAspSerThrSerSerProValHisSerSerAlaSerIleGlnAsnIle 262
QY 1722 ----- 1722
Db 263 LysThrThrSerAspLeuAlaSerThrProAspHisAsnGlyThrSerValThrThrThr 282
QY 1723 -----GGTGGCGAAGATATGTCAGACAC 1746
Db 283 SerSerAlaLeuGlySerAlaThrSerProAspHisSerGlyThrSerThrThrAsn 302
QY 1747 CTAGATGAATGATCTGTCTGACACTCTGCCCCATCTGAGGTACACAGCTCAGCGAA 1806
Db 303 SerSerGluSerValLeuAlaThrProValTyrSerSerMetProPheSerThrThr 322
QY 1807 TATGTTTCT-----GTCCCAGATCAT----- 1827
Db 323 LysValThrSerGlySerAlaIleIleProAspHisAsnGlySerSerValLeuProThr 342
QY 1828 -----TCTTGGAGGATACCACT----- 1845
Db 343 SerSerValLeuGlySerAlaThrSerLeuValTyrAsnThrSerAlaIleAlaThrThr 362
QY 1846 -----CCTGTCTCAGCTTTACAG 1863
Db 363 ProValSerAsnGlyThrGlnProSerValProSerGlnTyrProValSerProThrMet 382
QY 1864 TATATCACCACTAGTTCTATGACCATGCCCCAAGGCGGAGCTGTGTAGTGTCTTTC 1923
Db 383 AlaThrThrSerSerHisSerThrIleAlaSerSerTyrTyrSerThrValProPhe 402
QY 1924 AGT----- 1926
Db 403 SerThrPheSerSerAsnSerSerProGlnLeuSerValGlyValSerPhePheLeu 422
QY 1927 ---CTGGCTTGTCTAACATGGCTTCTCCACGACCTGTTCAACAGAGCTCTCTGGAG 1983
Db 423 PhePheTyrIleGlnAsnHisProPheAsnSerSerLeuGluAspProSerSerAsnTyr 442
QY 1984 TACCGAGCTCTGGAGCAACAATTCAACAGCTGTGTGTTCCATATCTACGATCCAATCTT 2043
Db 443 TyrGlnGluLeuLysArgAsnIleSerGlyLeuPheLeuGlnIlePheAsnGlyAspPhe 462
QY 2044 ACAGGATTTTAAGCAACTTGAATTACTTAACCTCAGAACCGGAGTGTGTTGTAATAGC 2103
Db 463 LeuGlyIleSerSerIleLys-----PheArgSerGlySerValValGluSer 479
QY 2104 AAAATCAAGCTTTGCTAAGTCTGTGCGGTATAACCTCACCAAGGCTGTGCGGGGTCTTG 2163
Db 480 ThrValValPheArgGluGlyThr-----PheSerAla 490
QY 2164 GAGGATTTTGTCTGTGTCGAGCCCAACACTCCATCTGGAAATAGACAGCTACTCTCTC 2223
```

```
Db 491 SerAspValLysSerGlnLeuIleGlnHis---LysLysGluAlaAspSerTyrAsnLeu 509
QY 2224 AACATTGAA-----CCAGCTGATCAAGCAGATCCC 2253
Db 510 ThrIleSerGluValLysValAsnGluMetGlnPheProSerAlaGlnSerArgPro 529
QY 2254 -----TGCAAGTCTCTGGCTCGGCG 2274
Db 530 GlyValProGlyTyrGlyIleAlaLeuLeuValValCysIleLeuValAlaLeuAla 549
QY 2275 -----GAATTTGCCCAATGCTGTAAGAACGAACGAGCTGAGGAAGCGGAG 2319
Db 550 IleValTyrPheLeuAlaLeuAlaValCys-----Gln 560
QY 2320 TGTGCGTGCACAAACCAGGATATACAGCCAGGCGGAGCTGGAC 2361
Db 561 CysArgArgLys-----SerTyrGlyGlnLeuAsp 570
```

Search completed: March 4, 2004, 18:48:16
Job time : 171.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 4, 2004, 19:38:06 ; Search time 161 Seconds
(without alignments)
13051.871 Million cell updates/sec

Title: US-10-007-270-1

Perfect score: 5875

Sequence: 1 taaccagaaggtatcct.....tactatatgacataatcaat 3330

Scoring table: BLOSUM62

| |
|---------------------------|
| Xgapop 10.0 , Xgapext 0.5 |
| Ygapop 10.0 , Ygapext 0.5 |
| Fgapop 6.0 , Fgapext 7.0 |
| Delop 6.0 , Delext 7.0 |

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-Q=/cgn2_1/USPTO_epool_p/US10007270/runat 04032004.160716.4957/app query.fasta.1.3527
-DB=SPTRMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=opt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10007270.cgn_1_181 @runat 04032004.160716.4957 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

SPTRMBL 25:*

```
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 4073 | 69.3 | 797 | 4 | 043686 homo sapien |

Q9GMS5 bos taurus
Q9ET31 mus musculus
Q9ET62 rattus norv
Q8LW8 mus musculus
Q9ES62 mus musculus
Q9J18 gallus gall
Q9BWI homo sapien
Q9UKK5 homo sapien
Q9BZV3 homo sapien
Q9L0Y3 mus musculus
P70628 rattus norv
Q80XK2 mus musculus
Q46616 macaca fasc
Q9CTP8 mus musculus
Q8C84 mus musculus
Q8LW8 mus musculus
Q9XN11 oryza sativ
Q9FJ35 arabidopsis
Q9K60 mus musculus
Q9NZ11 macaca mulla
Q8L0N0 plasmodium
Q87115 neurospora
Q8694 dictyosteli
Q9ES07 homo sapien
Q8L0N0 plasmodium
Q9L115 oryctolagus
Q9KTA5 vibrio chol
Q8IPJ3 drosophila
Q8L117 plasmodium
Q8L115 plasmodium
Q9K10 neisseria m
Q7Z466 homo sapien
Q8L112 plasmodium
Q9NFS3 drosophila
Q8IOG9 plasmodium
Q9VHJ7 mus musculus
Q8C1C0 mus musculus
Q8LX7 synechococc
Q8L0M9 plasmodium
Q9Z7K3 chlamydia p
Q8L119 plasmodium
Q9VQ8 drosophila
Q95YM2 procambatus

ALIGNMENTS

RESULT 1

043686
ID 043686 PRELIMINARY; PRT; 797 AA.
AC 043686; Q95094;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Interphotoreceptor matrix proteoglycan 150.
IMPGI OR IPML50.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA MEDLINE=20068045; PubMed=10601738;
RX Kuehn M.H., Hageman G.S.;
RT "Expression and characterization of the IPM 150 gene (IMPGI) product,
RT a novel human photoreceptor cell-associated chondroitin-sulfate
RL proteoglycan.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98358139; PubMed=9691169;

RA Feilbor U., Gehrig A., Sauer C.G., Marquardt A., Kohler M., Schmid M.,
 RT Weber B.H.;
 RT "Genomic organization and chromosomal localization of the
 RT interphosphoreceptor matrix proteoglycan-1 (IMP1) gene: a candidate for
 RT 6q-linked retinopathies.";
 RL Cytogenet. Cell Genet. 81:112-117(1998).
 DR EMBL; AF047492; AAC03789.2; -;
 DR EMBL; AF017776; AAC68835.1; JOINED.
 DR EMBL; AF017760; AAC68835.1; JOINED.
 DR EMBL; AF017761; AAC68835.1; JOINED.
 DR EMBL; AF017762; AAC68835.1; JOINED.
 DR EMBL; AF017763; AAC68835.1; JOINED.
 DR EMBL; AF017764; AAC68835.1; JOINED.
 DR EMBL; AF017765; AAC68835.1; JOINED.
 DR EMBL; AF017766; AAC68835.1; JOINED.
 DR EMBL; AF017767; AAC68835.1; JOINED.
 DR EMBL; AF017768; AAC68835.1; JOINED.
 DR EMBL; AF017769; AAC68835.1; JOINED.
 DR EMBL; AF017770; AAC68835.1; JOINED.
 DR EMBL; AF017771; AAC68835.1; JOINED.
 DR EMBL; AF017772; AAC68835.1; JOINED.
 DR EMBL; AF017773; AAC68835.1; JOINED.
 DR EMBL; AF017774; AAC68835.1; JOINED.
 DR EMBL; AF017775; AAC68835.1; JOINED.
 DR Genew; HGNC:6055; IMP1.
 DR GO; GO:0005578; C:extracellular matrix; TAS.
 DR GO; GO:0005201; E:extracellular matrix structural constituent; TAS.
 DR GO; GO:0007601; F:vision; TAS.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF01390; SEA; 2.
 DR SMART; SM00200; SEA; 2.
 DR PROSITE; PS00024; SEA; 1.
 KW Receptor.
 SQ SEQUENCE 797 AA; 89387 MW; D017ED090C78D521 CRC64;
 Alignment Scores:
 Pred. No.: 1.86e-289 Length: 797
 Score: 4073.00 Matches: 796
 Percent Similarity: 97.31% Conservative: 1
 Best Local Similarity: 97.19% Mismatches: 0
 Query Match: 69.33% Indels: 22
 DB: 4 Gaps: 1
 US-10-007-270-1 (1-3330) x 043686 (1-797)
 QY 131 ATGATATTGGAACACTAGAGAGCTATTGTTGTTTTCGATTTTTCACAGTTCAAGGA 190
 DB 1 MetTyrLeuGluThrArgAlaIlePheValPheThrIlePheLeuGlnValGlnGly 20
 QY 191 ACCAAGATATCTCCATTACATATACATATCCATTCTGAACTAAAGACATAGACATACCCCA 250
 DB 21 ThrLysAspIleSerIleAsnIleTyrHisSerGluThrLysAspIleAspAsnProPro 40
 QY 251 AGAATGAACAACTGAAAGTACTGAAAAATGTACAAAAATGTCAACTATGAGCAAGTA 310
 DB 41 ArgAsnGluThrThrGluSerThrGluLysMetTyrLysMetSerThrMetArgArgIle 60
 QY 311 TTCCATTTCGCAAGCATCGACAAAGATCCGATTTTCCCAACGGGGTTAAAGTC 370
 DB 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
 QY 371 TGTCACAGGAATCCATGAACAGATTTTAGACAGCTCTTCAAGCTTTATATAGATTGAGA 430
 DB 81 CysProGlnGluSerMetLysGlnIleLeuAspSerLeuGlnAlaTyrTyrArgLeuArg 100
 QY 431 GTGTGTGAGAGAGCATATGGGAAGCATATCGATCTTTCTGGATCGCATCCCTGACACA 490
 DB 101 ValCysGlnGlnAlaValTrpGluAlaTyrArgIlePheLeuAspArgIleProAspThr 120
 QY 491 GGGGAATATCAGAGCTGGGTCCAGCATCTCCAGAGAGAGACCTTTGCTCTTTGACATT 550
 DB 121 GlyGluTyrGlnAspTrpValSerIleCysGlnGlnGluThrPheCysLeuPheAspIle 140

QY 551 GGAAGAACTTCAGCAATTCACAGGAGACCTGGATCTTCTCCAGCAGAGATATAACAG 610
 DB 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuGlnGlnArgIleLysGln 160
 QY 611 AGAAGTTTCCCTGACAGAGAAAGATGAAATATCTCAGAGAGACATTCGGAGAGCTGGT 670
 DB 161 ArgSerPheProAspArgLysAspGluLleSerAlaGluLysThrLeuGluGluProGly 180
 QY 671 GAAACCATTTGCTATTTCAACAGCAATCTACATTTTCAAGACTTGGCAGATATTCTAAGAA 730
 DB 181 GluThrIleValIleSer--Thr----- 187
 QY 731 AACCTCAGAGAGCAATTCAGATGTTGCCAAGCTTCTCCTCCTTTCGGGCTTTCCTCTCA 790
 DB 188 -----AspValAlaAsnValSerLeuGlyProPheProLeuT 200
 QY 791 CTCTCATCACCCTCTCTCAATGAAATCTCGATATATACACTCAACACACCAAGATGC 850
 DB 200 hrProAspAspThrLeuLeuAsnGluLleLeuAspAsnThrLeuAsnAspThrLysMetP 220
 QY 851 CTACACAGAGAGAAAGAAAGATTCGCTGTGTGGAGAGCAGAGGTGGAGCTCAGCG 910
 DB 220 roThrThrGluArgGluThrGluPheAlaValLeuGluGluGlnArgValGluLeuSerV 240
 QY 911 TCTCTCTGTAAACAGAGCTTCAAGCAGAGCTCCCTGACTCCCTCCAGTCCCTTCCAGAT 970
 DB 240 aIserLeuValAsnGlnLysPheLysAlaGluLeuAlaAspSerGlnSerProTyrTyrG 260
 QY 971 AGAGCTCAGAGGAAAGTCCCACTTCAGATGCAAAAGATATTTAAGAAATCTTCAGAT 1030
 DB 260 InGluLeuAlaGlyLysSerGlnLeuGlnMetGlnLysIlePheLysLeuProGlyP 280
 QY 1031 TCAAAAAATCCATGCTGTTAGGATTTAGACCAAGAAAGAAAGATGGTCAAGCTCA 1090
 DB 280 helyLysIleHisValLeuGlyPheArgProLysLysGlyLysAspGlySerSerSert 300
 QY 1091 CAGAGATGCAACTTACGGCCATCTTTAAGAGACACAGTGACAGCAAGCAAAAGCCCTGAA 1150
 DB 300 hrGluMetGlnLeuThrAlaIlePheLysArgHisSerAlaGluAlaLysSerProAla 320
 QY 1151 GTGACCTCTGCTTTTGATTCCCAAAATTCGAAAGTGAGGAAAGTCTATCATGAAACA 1210
 DB 320 erAspLeuLeuSerPheAspSerAsnLysIleGluSerGluGluValTyrHisGlyThrM 340
 QY 1211 TGAGGAGAGCAAGCAAGCAAGCAAGTATCTCTCAGAGCTACAGACTCAAAAGCTGATCA 1270
 DB 340 elGluGluAspLysGlnProGluIleTyrLeuThrAlaThrAspLeuLysArgLeuLys 360
 QY 1271 GCAAGCACTAGAGGAGAAACAATCTTTGGATGTGGGACAAATTCAGTTTCATGATGAA 1330
 DB 360 erLysAlaLeuGluGluGlnSerLeuAspValGlyThrIleGlnPheThrAspGluI 380
 QY 1331 TTGCTGATCAGTCCAGGCTTTGGTCTTGACACCAATCAGAGCTGCCCATCTTTTG 1390
 DB 380 leAlaGlySerLeuProAlaPheGlyProAspThrGlnSerGluLeuProThrSerPheA 400
 QY 1391 CTGTATATAACAGAGATCTACTTTGAGTCCAGACTTCTCTCTCTGTTGAGACCCAGCTTG 1450
 DB 400 laValIleThrGluAspAlaThrLeuSerProGluLeuProValGluProGlnLeuG 420
 QY 1451 AGACAGTGGAGCGAGCAGCATGGTCTACCTGACACTTCTTGCTCTCCACCTGCTATGG 1510
 DB 420 luThrValAspGlyAlaGluHisGlyLeuProAspThrSerTrpSerProProAlaMetA 440
 QY 1511 CCTCTACCTCTCTGTCAGAGCTCCACCTTTCTTTATGGCATCAAGCATCTTCTCTCGA 1570
 DB 440 laSerThrSerLeuSerGluAlaProProPhePheMetAlaSerSerIlePheSerLeuT 460
 QY 1571 CTGATCAAGCAACACAGATACATGCGCACTCCAGACAAATGCTACTGTACAGGGCTCA 1630
 DB 460 hrAspGlnGlyThrThrAspThrMetAlaThrAspGlnThrMetLeuValProGlyLeuT 480
 QY 1631 CCATCCCAACAGTGATTTTCTGCAATCAGCAACTGGCTCTGGGAATTTTACATCCAC 1690


```

Db      480 hrllleProThrSerAspTyr-SerAlaIleSerGlnLeuAlaLeuGlyIleSerHisProp 500
QY      1691 CTGCATCTTCAGATCAGACGCGATCAAGTCAGAGTCGCGAAGATATGTCAGACACTAG 1750
Db      500 roalaSerSerAspSerArgSerAlaGlyGluAspMetValArgHisLeuA 520
QY      1751 ATGAATGGATCTGTGACACTCTCGCCCATCTGAGGTACAGAGCTCAGCGAATATG 1810
Db      520 spGluMetAspLeuSerAspThrProAlaProSerGluValProGluLeuSerGluTyrV 540
QY      1811 TTTCTCTCCAGATCATTTCTTGAGGATACACTCTCTGTCTCAGCTTTACAGTATATCA 1870
Db      540 alSerValProAspHisPheLeuGluAspThrThrProValSerAlaLeuGlnTyrIleT 560
QY      1871 CCCTAGTCTATACCATTTGCCATTTGCCAGGATACACTCTCTGTCTCAGCTTTACAGTATATCA 1870
Db      560 hrThrSerSerMetThrIleAlaProLysGlyArgGluLeuValValPhePheSerLeuA 580
QY      1931 GTTCTCTTAACATGGCTCTCTCAACGACCTGTTTCAACAGAGCTCTCTGAGTACCGAG 1990
Db      580 rgValAlaAsnMetAlaPheSerAsnAspLeuPheAsnLysSerSerLeuGluTyrArgA 600
QY      1991 CTCTGAGGACAAATTCACACAGCTGCTGTTCCATATCTACGATCCAAATCTTACAGAT 2050
Db      600 laeuglGlnGlnPheThrGlnLeuLeuValProTyrLeuArgSerAsnLeuThrGlyP 620
QY      2051 TTAAGCACTTGAATACTTAATCTTCAAGACGGAGTGTGATTTGTAATAGCAAAATGA 2110
Db      620 heLysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysMetL 640
QY      2111 AGTTTGTAGTCTGTGCGGTATAACTCTCAAGAGCTGTGACGGGCTCTTGGAGGATT 2170
Db      640 yepHeAlaLysSerValProTyrAsnLeuThrLysAlaValHisGlyValLeuGluAspP 660
QY      2171 TTCTGCTCTGTGAGCCCAACACTCTCAAGAGCTGTGACGGGCTCTTGGAGGATT 2230
Db      660 heArgSerAlaAlaGlnGlnLeuHisLeuGluIleAspSerTyrSerLeuAsnIleG 680
QY      2231 AACACGCTGATCAAGCAGATCCCTGCAAGTCTCTGCGCTGCGGCAATTTGCCATGTG 2290
Db      680 luProAlaAspGlnAlaAspProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysV 700
QY      2291 TAAGAAGCAACGGACTGAGGAGCGGAGTGTGCTGCAACACAGGATATGACAGCCAGG 2350
Db      700 alLysAsnGluArgThrGluGluAlaGluCysArgCysLysProGlyTyrAspSerGln 720
QY      2351 GGAGCTGGAGCGTCTGGAACCGGCTCTGTGCGCTGCGGCAATTTGCCATGTG 2410
Db      720 lySerLeuAspGlyLeuGluProGlyLeuCysGlyProGlyThrLysGluCysGluVal 740
QY      2411 TCCAGGAAAGGAGCTCCATGCGAGGTTCGAGATCACTCTGAAATCAAGCATACAAA 2470
Db      740 euGlnGlyLysGlyAlaProCysArgLeuProAspHisSerGluAsnGlnAlaTyrIleT 760
QY      2471 CTAGTGTAAAAGTTCCAAAATCAACAAATPAACAGGTAATCAGTAAAGAAATTCG 2530
Db      760 hrSerValLysLysPheGlnAsnGlnAsnAsnLysValIleSerLysArgAsnSerG 780
QY      2531 AATTACTGACCTGAGATATGAGAAATTTAACCATCAAGATTGGGAGGAAAT 2583
Db      780 luLeuLeuThrValGluTyrGluGluPheAsnHisGlnAspTrpGluGlyAsn 797

```

RESULT 2

```

Q9GMS5
ID Q9GMS5 PRELIMINARY; PRT; 794 AA.
AC Q9GMS5;
DT 01-WAR-2001 (T-EMBLrel. 16, Created)
DT 01-WAR-2001 (T-EMBLrel. 16, Last sequence update)
DE 01-WAR-2003 (T-EMBLrel. 23, Last annotation update)
DE MLGAPC core protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN NCBI_TaxID=9913;
RP SEQUENCE FROM N.A.
EX MEDLINE=20391374; PubMed=10937595;
RA Uehara F., Ooba N., Ozawa M.;
RT "Isolation and Characterization of Mucinlike Glycoprotein Associated
RT with Photoreceptor Cells.";
RL Invest. Ophthalmol. Vis. Sci. 41:2759-2765 (2000).
DR EMBL; AB047844; BAB12254.1; -.
DR InterPro; IPR006210; IEGF.
DR PFAM; PF01390; SEA; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00200; SEA; 2.
DR PROSITE; PS00024; SEA; 1.
SQ SEQUENCE 794 AA; 89377 MW; 4965FD322FA95841 CRC64;

```

Alignment Scores:

```

Pred. No.: 3,54e-190 Length: 794
Score: 2717.00 Matches: 558
Percent Similarity: 76.90% Conservative: 81
Best Local Similarity: 67.15% Mismatches: 142
Query Match: 46.25% Indels: 51
DB: 6 Gaps: 8

```

US-10-007-270-1 (1-3330) x Q9GMS5 (1-794)

```

QY 131 ATGATTTGGAACTAGACAGCTATTTTGTGATTTTCTCCAGTTCAGGA 190
Db 1 MetHisLeuGluAlaAlaArgValIlePhePheLeuTyrIlePheLeuGlnValGlnGly 20
QY 191 ACCAAGATATCTCCATTAACATATACCATCTCGAACTTGAACCTAAGACATAGCAATCCCCA 250
Db 21 IleLysAspLeuSerIleLysIleTyrGlySerGluIleLysAspIleAspAsnAlaPro 40
QY 251 AGAATTAAGAACCACTGAAGTACTGAAATAATGTCAAAATGTCAAAATGTCACTATGACGGAATA 310
Db 41 ArgThrGluAlaThrLysAsnThrAlaLysThrLysValSerThrMetArgArgIle 60
QY 311 TTGATTTGGCAAGCATCATCAAAAGATCCGCAATTTTCCCAAGGGGGTTAAAGTC 370
Db 61 PheAspLeuAlaLysHisArgThrLysArgSerAlaPhePheProThrGlyValLysVal 80
QY 371 TGTCCACAGGAATCCATGAACAGATTTTACAGCTCTTCAAGCTTTATATAGATTGAGA 430
Db 81 CysProGlnGluSerMetGluGlnIleLeuAlaSerLeuGlnAlaTyrTyrArgLeuArg 100
QY 431 GTGTCTCAGGAACAGTATGGGAAGCATATCGGATCTTTCTGGATCGCATCCCTGCACACA 490
Db 101 ValCysGlnGluAlaValTyrGluAlaTyrArgIlePheLeuAspArgLeuProGluPro 120
QY 491 GGGGAATATCAGGACTGGCTGAGCATCTGCCAGCAGGAGACCTTCTGCTCTTTGACATT 550
Db 121 GlyGluTyrGlnAspTyrValSerValCysGlnGlnGlnThrPheCysLeuPheAspIle 140
QY 551 GGAATAAATCTCAGCAATTTCCAGGAGCACCTGGATCTTCCAGCAGGAGATATAACAG 610
Db 141 GlyLysAsnPheSerAsnSerGlnGluHisLeuAspLeuLeuGlnGlnArgMetLysGln 160
QY 611 AGAATTTCCCTGACAGAAAAGATGAATATCTCGACAGAGAAGACATTTGGAGAGCCTGGT 670
Db 161 ArgAsnPheLeuGluArgLysAspGluValValThrLysGlnThrLeuGlyGluLeuGly 180
QY 671 GAAACCATGTTTCATTTCAACAGCAATCTACATTTCAAAAGACTTGGGCGAGTATTTCAAGAA 730
Db 181 GlnThrProGlyLeuGln- 186
QY 731 AACCCCTCAGAGAGCAAAATTCAGATGTTTGCACAGCTCCTACCTGGGCTTTCCCTCTCA 790
Db 187 -----GlnThrLeuProValSerHisProGlyProCysLeuSer 199

```

```

QY 791 CTCTGATGACACCTCTCTCAATGAATTTCTCGATAATACATCTCAACGACCAACAGATG- 849
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 LeuProMetThrThrAlaGlnArgAenProGlnLeuHisProSerArgThrProArgVal 219
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 850 CCTACACAAAGAGAGAAACAGAAATTCGCTGTGTGGAG-----GAGCAGAGG 897
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 ProThrArgGluArgLysIleGluPheThrAspAlaAlaGluAspAlaLeuGluGlnIuys 239
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 898 GTGGAGCTCAGGCTCTCTGTGTAACACAGAGTTCAAGCGCAGAGCTCGTCACTCCAG 957
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 ValGluLeuSerIleSerLeuAlaAenGlnIysPheLysSerGluLeuAspAenSerGln 259
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 958 TCCCATATTACAGAGCTAGCAGAGAGTCCCAAGTCCAGATTCAGATCAAAAGATATTAA 1017
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 SerProTyrTyrLeuGluValAlaAlaLysSerGlnLeuMetGlnLysIlePheLys 279
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1018 AAATCTTCAGGATCAAAAATCCATGTGTAGGATTTAGACCAAGAAAGAAAAGAT 1077
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 LysLeuProGlyPheLysGluIleValSerGlyPheArgProLysLysGluArgAsp 299
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1078 GGCTCAAGCTCCACAGAGATGCACTTACGGCCATCTTTAAGACACAGTGCAGAGCA 1137
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 GlyThrSerSerThrGluMetGlnLeuThrAlaIlePheLysLysGlyLysAlaGluAla 319
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1138 AAAAGCCCTGCAAGTGACCTCTCTGTTTTCATTCCAACAAATTTGAAGTGGAGAGTC 1197
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 LysSerProAlaSerAspProLeuSerLeuAspSerAsnLysIleGluSerGluGlyAsp 339
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1198 TATCATGACATGGAGGAGGACCAAGCAACCAAGAAATCTATCTCACAGCTACAGACCTC 1257
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 ProArgIleThrGluGluGluLysGlnArgGluLeuTyrProThrAlaSerGluLeu 359
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1258 AAAAGCTGATCACAAAGCACTAGAGGAGAGCAATCTTGGATGTGGGAGCAATTCAG 1317
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 ArgLysLeuIleSerArgAlaLeuGluGluAspGlnSerLeuAspValGlyThrIleGln 379
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1318 TTCACGTGAATTTGCTGGATCACTGCCAGCTTGTGCTGTGACCCCAATCAGAGCTG 1377
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 PheThrAspGluIleValGlySerLeuProSerLeuAspProAspThrGlnLeuValLeu 399
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1378 CCCACATCTTTGCTGTTATACAGAGATGCTACTTTGAGTCCAGAACTTCTCTCTCTT 1437
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 ProThrLeuLeuThrAspIleThrLysAspAlaThrLeuSerProGluLeuProLeuGly 419
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1438 GAACCCAGCTTCAGACAGTGCAGCGAGCAGACATGTCTACTGACACTTCT----- 1491
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 GlnProArgLeuGluThrValAspArgAlaGlyHisSerProGlyAlaSerProThr 439
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1492 -----TGGTCTCCACCTCTATGGCTCTACTCTCTCTGTCAGAGCTCCACTTCTTT 1545
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 AspGlyTyrSerProAlaMetThrSerThrSerLeuSerGluThrLeuProPhePhe 459
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1546 ATGGCATCAACGATCTCTCTGATCAAGGACCAACAGATCAATGACCACTGAC 1605
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 ThrAlaSerSerValPheProGlnThrAspGlnSerAlaThrAspIleMetSerIleAsp 479
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1606 CAGACATGTAGTACAGGGCTCACCATCCACAGCTGATTTCTTCATCAGCA 1665
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 GlnThrValLeuLeuProArgLeuThrValProThrAspAspTyrSerAlaIleSerPro 499
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1666 CTGGCTTCGGAATTTACATCCACCTGCATCTTCAGATGACAGCGAGTCAAGTCAGGT 1725
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500 LeuValProGluIleSerHisLeuProThrSerSerGluAspTrpLeuSerSerSer--- 518
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1726 GCGCAAGATGTGTACAGACTAGATGAATGGATCTGTCTGACACTCTCTGCCCATCT 1785
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 SerGlnAspThrMetGluTyrLeuAspGlyValAspLeuThrLysThrProThrSerSer 538
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1786 GAGGTACAGAGCTCAGC-----GAATATGTTTCTCTCCAGATCATTC 1830
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 GluGlyProArgAsnSerValGlyMetPheProAlaIlePheIle-----Phe 554
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1831 TTGAGGAGTACCACCTCTCTCAGCTTTTACAGTATATCACCACCTAGTTCTATGACCAT 1890
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 555 LeuGluAenIleThrProAspProGlyLeuArgTyrIleThrThrSerAlaMetThrVal 574
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 GCCCCCAGGGCGGAGAGCTGTAGTGTCTTCAGTCTGCGTGTCTTCAATCATGCGCTTC 1950
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 AlaAlaArgGlyArgGluLeuValPhePheSerLeuArgValAlaAenValProPhe 594
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1951 TCCAAACACCTGTCAACAGAGCTCTCGGAGTACCGAGCTCTGAGCAGCAATTCACA 2010
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 SerThrAspLeuPheAenLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThr 614
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2011 CAGCTGCTGTGTTCCATATCTACCATCTTACAGATTTAAGCACTTGAATTAATTA 2070
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 GlnLeuLeuValProAenLeuArgSerAenLeuThrGlyPheLysGlnLeuGluIleLeu 634
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2071 AACTTCAGAAACGGAGTGTGATTGGAATGCAAAATGAAGTTTGTCTTAAGTCTGTGCCG 2130
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 AsnPheArgAsnGlySerValIleValAenSerLysValArgPheAlaLysSerValPro 654
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2131 TATAACTCTCAACAGGCTGTGCAGGGGTCTTGGAGGATTTTCTTCTGCTGCGAGCCAA 2190
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 TyrAsnLeuThrLysAlaValArgGlyValLeuGluAspPheArgSerThrAlaAlaGln 674
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2191 CAATCTCATCTGGAATAGACACTCTCTCAACATTTGAACCACTGATCAAGCAGAT 2250
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 GlnLeuAspLeuGluIleAspSerTyrSerLeuAspValGluProAlaAspGlnAlaAsp 694
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2251 CCTGTCAAGTCTTCTGCGCTCGGCGCAATTTGCCCATGTGTAAAGAACAGGACTGAG 2310
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 ProCysLysPheLeuAlaCysGlyGluPheAlaGlnCysValArgAenGluTrpThrGlu 714
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2311 GAAGCGGAGTGTCTGCAAAACAGGATATCAGACCCAGGAGGCTCGACGCTCTGGA 2370
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 GluAlaGluCysArgCysArgSerGly-----ThrGlnAlaLeuValLeuProIleGlu 732
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2371 CGAGGCTCTGTGCGCTTGGCCCTTGGCCCAAGGAATGCGAGTCTCTCCAGGGAAGGAGTCCA 2430
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 -----AspCysGluAspIleProGlyLysGlyThrPro 743
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2431 TGCAGGTGCCAGATCACTCTGAAATCAAGCATACAAACACTAGTGTAAAGATTCCAA 2490
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 CysArgSerLeuAspGlnSerLysAenGlnValTyrGluProGlyValLysPheGln 763
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2491 AATCAACAAAATAACAGGTATATCAGTAAAGAAATTCAGTAACTACTGACCGTAGATAT 2550
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 764 ArgGlnGlnAspAenLysValThrMetLysArgLysPheGluLeuLeuThrIleGlyTyr 783
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2551 GAAGAAATTTAACCATCAAGATTTGGGAAGGAAAT 2583
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 GluGluPheAsnTyrGlnAspTrpGluGlyAsn 794
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 3

```

Q9ET31
ID Q9ET31 PRELIMINARY; PRT; 798 AA.
AC Q9ET31;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE IMP 150 proteoglycan.
GN IMPG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RT Kuehn M.H., Wietcki D.T., Hageman G.S.;
RT "Molecular characterization of the murine orthologue of the human
RT retinal proteoglycan IMP 150.";
RL Mol. Vis. 6:0-0(2000)
DR EMBL; AF266478; AAC00796.1; -.
DR MGD; MGI:1926876; Impg1.
DR InterPro; IPR000082; SEA_domain.

```

DR Pfam; PF01390; SEA; 2.
 DR SMART; SM00200; SEA; 2.
 SQ SEQUENCE 798 AA; 89542 MW; 170D2CB94ACD76A2 CRC64;

Alignment Scores:

Pred. No.: 1.28e-174 Length: 798
 Score: 2504.50 Matches: 527
 Percent Similarity: 73.13% Conservative: 80
 Best Local Similarity: 63.49% Mismatches: 180
 Query Match: 42.63% Indels: 43
 DB: 11 Gaps: 10

US-10-007-270-1 (1-3330) x Q9ET31 (1-798)

```

QY 131 ATGTATTTTGGAACTAGACAGCTATTTTGTGTTTGGATTTTCTCCAGTTCAGGA 190
Db      :::::
QY 1 MeAsnPhenGluLeuHisAlaIlePheValPheGlyIlePheLeuGlnValGlnGly 20
Db      :::::
QY 191 ACCAAGATATCTCCATTAACATATACCATCTGAAACTAAAGACATAGACATCCCCA 250
Db      :::::
QY 21 IleLysAspThrSerIleLysIlePheSerSerGluLelLysAsnIleAspLysThrPro 40
Db      :::::
QY 251 AGAATGAAACAACTGAAAGTACTGAAAAAATGTACAAAATGTCAAAATGATCACTATGACAGCAATA 310
Db      :::::
QY 41 ArgIleGluThrIleGluSerThrSerThrValHisIleValSerThrMetLysArgIle 60
Db      :::::
QY 311 TTCGATTTGGCAACATCGAACAAAGATCCGCATTTTCCCAACGGGGGTTAAAGTC 370
Db      :::::
QY 61 PheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro---AlaAlaAsnIle 79
Db      :::::
QY 371 TGTCACAGGAATCCATGAACAGATTTTAGACAGCTTCAAGCTTATATAGATGAGA 430
Db      :::::
QY 80 CysProGlnGluSerLeuArgGlnIleLeuAlaSerLeuGlnGluLysArgLeuArg 99
Db      :::::
QY 431 GTGTCTCAGGAGCAGTATGGAGCATATCGGATCTTTCTGGATCGCATCCCTGCACACA 490
Db      :::::
QY 100 ValCysGlnGluValValTrpGluAlaTrpArgIlePheLeuAspArgIleProAspThr 119
Db      :::::
QY 491 GGGGAATATCAGGACTGGGTGAGCATCTGCCACGACGAGACCTTCTGCTTTTGCACATT 550
Db      :::::
QY 120 GluGluTrpGlnAspTrpValSerLeuGlnGlnGluLeuThrPheCysLeuPheAspIle 139
Db      :::::
QY 551 GGAATAAACTTACGCAATTCACGAGGAGCCTCGATCTTCTCCAGCAGAGATTAACAG 610
Db      :::::
QY 140 GlyLysAsnPheSerAsnSerGlnGlnHisLeuAspLeuGlnGlnArgIleLysGln 159
Db      :::::
QY 611 AGAAGTTTCCCTGACAGAAAGATGAATATTCGACAGAGAGACATTTGGAGAGCGCTGGT 670
Db      :::::
QY 160 ArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeuGluAlaProThr 179
Db      :::::
QY 671 GAAACCATTTGTCATTTCAACAGCAATCTACATTTCAAGACTTGGGAGATTTCTAAGAA 730
Db      :::::
QY 180 GluAlaProValPro---Thr----- 186
Db      :::::
QY 731 AACCTCAGAAGAGCAAAATTCAGATGTTGCCAAGCTCTCACTTGGGCGCTTCCCTCTCA 790
Db      :::::
QY 187 -----AspValSerArgMetSerLeuGlyProPheProLeuP 199
Db      :::::
QY 791 CTCCTGATGACACCTCTCCTCAATGAATTCCTGATAATCACTCAACGACACCAAGATGC 850
Db      :::::
QY 199 roSerAspAspThrAspLeuLysGlnIleLeuSerValThrLeuLysAspIleGlnLysP 219
Db      :::::
QY 851 CTCACACAGAAAGAGAAACA-----GAAATCGCTGTGTGGAGGAGC 892
Db      :::::
QY 219 roThrThrGluSerLysThrGluProIleHisValSerGluPheSer-----SerGluG 237
Db      :::::
QY 893 AGAGGTGGAGCTCAGCGTCTCTCTGTAACACGAGAGTTCAAGGACAGAGCTCGCTCACT 952
Db      :::::
QY 237 LuLysValGluPheSerIleSerLeuProAsnHisArgPheLysAlaGluLeuThrAsnS 257
Db      :::::
QY 953 CCCAGTCCCATTATTCAGGAGCTAGCGAAAGTCCCAACTTCCAGATGCAAAAGATAT 1012
Db      :::::
QY 257 erGlySerProTyrrGlnGlnLeuValGlyGlnSerGlnLeuGlnLeuGlnLysIleP 277
Db      :::::

```

```

QY 1013 TTAAGAACTTCCAGGATTCAAAAAATCCATGTGTAGGATTTAGACCAAAAGAAAGAA 1072
Db      :::::
QY 277 heLysLysLeuProGlyPheGlyLysValLeuArgValLeuGlyPheArgProLysLysGluG 297
Db      :::::
QY 1073 AAGATGGCTCAAGCTTCCACAGAGATGCAACTTACGCGCATCTTTAGAGACACAGTGGAG 1132
Db      :::::
QY 297 LuAspGlySerSerThrGluIleGlnLeuMetAlaIlePheLysArgAspHisAlaG 317
Db      :::::
QY 1133 AAGCAAAAGCCCTGCAAGTCACTCTCTCTTTTGAATTCACAAATTTGAAAGTCAAGG 1192
Db      :::::
QY 317 LuAlaLysSerProAspSerHisLeuSerLeuAspSerAsnLysIleGluSerGluA 337
Db      :::::
QY 1193 AAGCTATATCATGAAACCATGAGGAGGACAGCAACACAGAAATCTATCTCAGCTACAG 1252
Db      :::::
QY 337 rgIleHisGlyValIle---GluAspLysGlnProGluThrTrpLeuThrAlaThrA 356
Db      :::::
QY 1253 ACCTCAAAAGGCTGATCAGCAAAAGCCTAGAGGAGAAACAATCTTTGGATGTGGGACAA 1312
Db      :::::
QY 356 spLeuLysLysLeuIleIleGlnLeuLeuAspGlyAspLeuSerLeuValGluGlyLysI 376
Db      :::::
QY 1313 TTCAGTTCACTGATGAATTTGGTGGATCACTGCCAGGCTTTTGGTCTCTGACACCAATCAG 1372
Db      :::::
QY 376 leProPheGlyAspGluValThrGlyThrLeu-----PheArgProValThrGluProA 394
Db      :::::
QY 1373 AGCTGCCACATCTTTTGTCTGTATACAGAGGATGCTACTTTAGTCCAGAACTTCCTC 1432
Db      :::::
QY 394 spLeuProLysProLeuAlaAspValThrGluAspAlaThrLeuSerProGluLeuProP 414
Db      :::::
QY 1433 CTGTGCAACCCGCTTGAGACAGTGGAGCGGACGAGCATGCTCTACT----- 1482
Db      :::::
QY 414 heValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuProGlyMetSerS 434
Db      :::::
QY 1483 -----CACACTTCTTGTCTCCACCTGCTATGCTCTTACCTCCCTGTCAGAACTCCAC 1537
Db      :::::
QY 434 erLysAspSerSerTrpSerProValSerAlaSerIleSerArgSerGluAsnLeuP 454
Db      :::::
QY 1538 CTTTCTTTATGGCATCAAGCATCTTCTCTGCTGATCAAGGCACACGACATCAATGG 1597
Db      :::::
QY 454 roSerPhe---ThrProSerIlePheSerLeuAspAlaGlnSerProProLeuMetT 473
Db      :::::
QY 1598 CACATGACACCAAGTCTAGTACCAAGGCTTCCACCATCCCAACAGTATTATTCTGCAA 1657
Db      :::::
QY 473 hrThrGlyProThrAlaLeuIleProLysProThrLeuProThrIleAspTrpSerThrI 493
Db      :::::
QY 1658 TCAGCCCACTGGCTCTGGGAATTTCAATCCACTCTGATCTTCAGATGACAGCCGACAA 1717
Db      :::::
QY 493 leArgGlnLeuProLeuGluSerSerHisTrpProAlaSerSerSerAspArgGluLeuI 513
Db      :::::
QY 1718 GTGCAAGTGGCGAAGATATGTCAGACACATAGATGAATGATCTCTCTGACACTCTTG 1777
Db      :::::
QY 513 leThrSerSerHisAspThrIleArgAspLeuAspGlyMetAspValSerAspThrProA 533
Db      :::::
QY 1778 CCGCATCTGAGGTACCAAGCTTCCAGCAATATATGTTTCTCCAGATCATTTCTTGGAGG 1837
Db      :::::
QY 533 laLeuSerGluIleSerGluLeuSerGlyTyraAspSerAlaSerGlyGlnPheLeuGluM 553
Db      :::::
QY 1838 ATACCACTCTGCTCAGCTTTTACAGTATATACCACTAGTCTTATGATGACCATGCCCCA 1897
Db      :::::
QY 553 etThrThrProIleProThrValArgPheIleThrThrSerSerGluThrIleAlaThrL 573
Db      :::::
QY 1898 AGGCGCAGAGCTGGTAGTGTCTTCTCAGTCTGGTGTCTTAACATGCTCTTCTCCACG 1957
Db      :::::
QY 573 ydGlyGlnGluLeuValPhePheSerLeuArgValAlaAsnMetProPheSerTrpA 593
Db      :::::
QY 1958 ACCTGTTCAACAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACACAGTGGC 2017
Db      :::::
QY 593 spLeuPheAsnLysSerLeuGluLysGlnAlaLeuGluGlnArgPheThrAspLeuL 613
Db      :::::
QY 2018 TGGTTCATATCTACGATCCCACTTACAGGATTTTACGAACTTGAATACTTACTTCA 2077
Db      :::::
QY 613 euValProTyrrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluIleLeuSerPheA 633
Db      :::::

```



```
QY 1241 TCACAGCTACAGACTCACAAGGCTGATCAGCAAGACCTAGAGAGAGACAATCTTTGG 1300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 euThraAlaAlaAspLeuYsLeuIleAlaArgLeuLeuAspGlnProLeuV 375
QY 1301 ATGTGGGGAACAATTCAGTTCACCTGATGAATTCGTGATCACTGCGAGCTTTGGCTCG 1360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
375 alGlyGlyThrValProPheSerAspGluVal-----ThrGluProLeuPheArgProV 393
QY 1361 ACACCCATCAGAGCTGCCCAACATCTTTGCTGTATATAACAGAGGATCTACTTTGAGTC 1420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
393 alThrGlnSerGluLeuProLysProLeuThrAspValThrGluAspValThrLeuSerP 413
QY 1421 CAGAACTCTCTCTGTTGAACCCAGCTTGAGACAGTGCAGCGGACGAGCATGTCTTAC 1480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 roGluLeuPheSerGluProArgLeuGlnSerValAspIleTyrGlyProTyrLeuP 433
QY 1481 CTGACACTTCTGCTTCCACCTGCTATGCGCTCTACCTCC---CTGTGAGAGCTCCAC 1537
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 roAspSerSerTyrPheArgProValThrAlaSerThrSerGlyValGlyAsnLeuProS 453
QY 1538 CTTTCTTATGGCATCAGCATCTTCTCTGACTGATCAAGGCACACAGATACAATGG 1597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
453 erPhe-----ThrProSerIlePheAlaLeuAspGlnSerSerProProLeuMeta 471
QY 1598 CCACTGACCAAGCAATGCTAGTACCAGGGCTCACCATCCCAACAGTATTTCTGCA 1657
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
471 laThrGlyProThrAlaPheIleProThrLeuThrLeuProIleSerAspTyrSerThrV 491
QY 1658 TCAGCCAACTGCGTCTGGGAATTCACAT---CCACTGCTCATCTTCAGATGACGCCGAT 1714
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
491 alArgGlnTyrProLeuGluValSerHisTyrProGluSerSerSerAspArgGluLeuS 511
QY 1715 CAAGTGCAGGTGCGGAGATATGCTCAGACACTAGATGAATGATCTGTCTGACACTC 1774
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
511 erThrThrSerSerHisAspThrIleArgAspLeuAspGluMetAspValSerAspThrP 531
QY 1775 CTGCCCCATCTGAGGTACAGAGCTCAGCGATATGTTCTGTCCCAAGATATTTCTTGG 1834
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
531 roAlaLeuSerGluIleAlaGluLeuSerGlyTyrAspSerAlaProAspArgPheLeuG 551
QY 1835 AGGATACACTCTCTGCTCAGCTTTACAGTATATACCACTAGTCTTATGACCAATCCCC 1894
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 luMetThrThrProIleProThrLeuGlnTyrValThr***SerSerGluThrIleAla 571
QY 1895 CCAAGGCGCAGAGCTGCTAGTGTCTTCACTGCTCGTGTGCTAACATGCGCTTCTCCA 1954
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
571 laLysGlyHisGluLeuValValPhePheSerLeuArgValAlaAsnMetProPheSerT 591
QY 1955 ACGACTGTTCACCAAGAGCTCTCTGAGTACCGAGCTCTGGAGCAACAATTCACAGC 2014
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
591 yrAspLeuPheAsnLysSerSerLeuGluTyrGlnAlaLeuGluGlnArgPheThrAspL 611
QY 2015 TGCTGTGTTCCATATCTAGATCCAACTTTACAGGATTAAGCAACTTGAATACTTAACT 2074
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
611 euLeuValProTyrLeuArgSerAsnLeuThrGlyPheLysGlnLeuGluIleLeuSerP 631
QY 2075 TCAGAAACGGAGTGTGATGTGATAGCAAAATGAAGTTGCTAGTCTGCTGCGGTATA 2134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
631 heArgAsnGlySerValIleValAsnSerLysValArgPheAlaLysAlaValProTyrA 651
QY 2135 ACCTCACCAAGGCTGTGACGGGCTCTGGAGGATTTTGGTCTGCTGCGGCCCAACAC 2194
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 snLeuThrGlnAlaValArgGlyValLeuGluAspLeuArgSerThrAlaAlaGlnGluL 671
QY 2195 TCATCTGGAATAGACAGCTACTCTCTCAACATTGACCACTGATCAAGCAGATCCCT 2254
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
671 euAsnLeuGluIleGluSerTyrSerLeuAspIleGluProAlaAspGlnAlaAspProC 691
QY 2255 GCAAGTTCCTGCGTGGCGGAATTTGCCAATGTGTAAAGAACCAAGCGAGCTGAGAG 2314
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
691 ysLysPheLeuAspCysGlyLysPheAlaGlnCysIleLysAsnGluLeuThrGluGluA 711
```

```
QY 2315 CGAGTGTCTGCTGCAACACAGGATATGATCAGCCAGGGGCGCTTGAGGCTCTGGAACAG 2374
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 laGluCysArgCysArgGlnIleHisGluSerHisGlyThrLeuGluTyrGlnGluLeuA 731
QY 2375 GCCTCTGTGCTGCTGCGCAAAAGAAATGCGAGGTCTCTCCAGGGAAGGAGGCTCCATGCA 2434
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
731 snLeuCysProProGly---LysThrCysGluAlaSerGlnGlyGlnAlaThrProCysA 750
QY 2435 GTTTGGCCAGTCACTCTGAAATCAAGCATACAAACTAGTGTAAAAGTTCCAAAATC 2494
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
750 tGProProAspHisSerThrAsnGlnAlaArgGlnProSerValLysLysLeuGlnArgG 770
QY 2495 AACAAATTAACAGTAAATCAGTAAAGAAATTTCTGAATTAATCTGACCGTACAGATACAG 2554
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
770 lngln---AsnLysValValLysLysArgAsnSerGluLeuSerAlaThrAspPheGluG 789
QY 2555 AATTAAACCATCAGATTGGGAAGGAAAT 2583
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
789 luLeuAspAspGlnAspTyrGluGlyAsn 798

RESULT 5
Q8R1W8
ID Q8R1W8 PRELIMINARY; PRT; 798 AA.
AC Q8R1W8;
DC 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 24, Last annotation update)
DE Interphotoreceptor matrix proteoglycan 1.
GN IMPG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022970; AM22970.1; -.
DR MGD; MGI:1926876; Impg1.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 2.
DR SMART; SM00200; SEA; 2.
KW Receptor.
SQ SEQUENCE 798 AA; 89474 MW; E875FB587389D4C CRC64;

Alignment Scores:
Pred. No.: 1.8e-174 Length: 798
Score: 2502.50 Matches: 527
Percent Similarity: 73.13% Conservative: 80
Best Local Similarity: 63.49% Mismatches: 180
Query Match: 42.60% Indels: 43
DB: 11 Gaps: 10

US-10-007-270-1 (1-3330) x Q8R1W8 (1-798)
QY 131 ATGATTATGGAACTAGAGAGCTATTTGTTTGGATTTTCTCCAGTTCAAGGA 190
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetAsnLeuGlnIleLysHisAlaIlePheValLeuGlyIlePheLeuGlnValGlnGly 20
QY 191 ACCAAGATATCTCCATTAACATATACCATCTCTGAACTAAAGACATAGACAATCCCCCA 250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
21 IleLysAspThrSerIleLysIlePheSerSerGluIleLysAsnIleAspLysThrPro 40
QY 251 AGAAATGAACCACTGAAGTACTGAAAAATGTACAAAATGTCAAACTATGAGACGAATA 310
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 ArgIleGluThrIleGluSerThrThrValHisLysValSerThrThrWeLysArgIle 60
QY 311 TTCATTTCGCAAGACATCCACAAAAGATCCCATTTTCCCAACGGGGGTTAAAGTC 370
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 PheAspLeuProLysLeuArgThrThrLysArgSerAlaLeuPhePro---AlaAlaAsnIle 79
```


Db 770 InGlnAenLysValValluLysArgAsnSerLysLeuSerAlaIleGlyPheGluGluP 790
 QY 2558 TTAACCATCAAGATTGGAGGAAAT 2583
 Db 790 heGluaspGlnaspTrpGluGlyAsn 798
 RESULT 6
 Q9ES62 PRELIMINARY; PRT; 742 AA.
 AC Q9ES62;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Sialoprotein associated with cones and rods SPACK.
 GN IMPG1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lee J.W., Chen Q., Rayhorn M.E., Shadrach K.G., Crabb J.W.,
 RA Rodriguez I.R., Hollyfield J.G.;
 RT "SPACK In the interphotoreceptor matrix of the mouse retina: molecular
 RT and biochemical characterization";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF229929; AAC32162.1; -.
 DR MGD; MGI:1926876; Impgl.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF01390; SEA; 2.
 DR SMART; SM00200; SEA; 2.
 SQ SEQUENCE 742 AA; 83177 MW; 58CAOCFE0FOA9B97 CRC64;

Alignment Scores:

Pred. No.: 7.66e-163 Length: 742
 Score: 2343.50 Matches: 494
 Percent Similarity: 73.13% Conservative: 72
 Best Local Similarity: 63.82% Mismatches: 165
 Query Match: 39.89% Indels: 43
 DB: 11 Gaps: 10

US-10-007-270-1 (1-3330) x Q9ES62 (1-742)

QY 299 ATGAGCAATATTCGATTGGCAAGATCGAACAAGAGATCCGATTTTCGCAAG 358
 Db 1 MetLysArgilePheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro--- 19
 QY 359 GGGTTAAAGTCGTCCACAGATCCATCAACAGATTTTACAGACTCTTCAAGCTTAT 418
 Db 20 AlaAlaAsnileCysProGlnGluSerLeuArgGlnileLeuAlaSerLeuGlnGluTyr 39
 QY 419 TATAGATTGAGAGTGTGCGAGAGCAGTATGCGAGCATATCGGATCTTTCTGATGCG 478
 Db 40 TyrArgLeuArgValCysGlnGluValValTrpGluAlaTrpArgilePheLeuAspArg 59
 QY 479 ATCCCTGACAGGGGATATCAGGACTGGGTGAGCATCTGCGAGAGAGACCTTCTGCG 538
 Db 60 IleProAspThrGluGluTyrGlnAspTrpValSerLeuCysGlnLysGluThrPheCys 79
 QY 539 CTCTTTGACATTTGGAATAAATCTTCAGCAATTCAGAGACATCCGATCTTCTCCAGCAG 598
 Db 80 LeuPheAspilleGlyLysAsnPheSerAsnSerGlnGlnileLeuAspLeuLeuGlnGln 99
 QY 599 AGATAAAGACAGAGATTCCCTGACAGAAAGATGAATATCTGCGAGAGAGACATTTG 658
 Db 100 ArgileGlnGlnArgSerPheProGlyArgLysAspGluThrAlaSerMetGluThrLeu 119
 QY 659 GGAGAGCCTGTGTGAACCAATGTTCATTTCAACAGCAATCTACATTTCAAGACTTGGCA 718
 Db 120 GluAlaProThrGluAlaProValPro--Thr----- 130
 QY 719 GTATTCTTAAGAAAACCTTCAGAGAGACAAATTCAGATGTTGCCAATCTCTACTTGGGC 778
 Db 1 MetLysArgilePheAspLeuProLysLeuArgThrLysArgSerAlaLeuPhePro--- 19

Db 131 -----AspValSerArgMetSerLeuGlyP 139
 QY 779 CTTTCCCTCTCACTCTCTGATGACACCCCTCTCTCATGAAATTCGATATACATCAACG 838
 Db 139 roPheProLeuProSerAspAspThrAspLeuLysGluileLeuSerValThrLeuLysA 159
 QY 839 ACACCAAGATGCTTACACAGAAAGAGAAACA-----CAATTCCTCTG 880
 Db 159 splleGlnLysProThrThrGluSerileThrGluProIleHisValSerGluPheSer- 178
 QY 881 TGTGGAGGAGCAGAGAGGTGAGCTCAGCGTCTCTCTGGTAAACACAGAAAGTTCAAGCAG 940
 Db 179 -----SerGluLulysValGluPheSerileSerLeuProAsnHisArgPheLysAlaG 197
 QY 941 AGTCGCTGACTCCCTCCCTCCCTATATACAGAGAGCTAGCAGAAAGTCCCACTTCAGA 1000
 Db 197 lueuThrAsnSerGlySerProTyrGlnGluLeuValGlnSerGlnLeuGlnL 217
 QY 1001 TGCAAAAGATATTTAAGAAATTTCCAGGATTCAAAATAATCCATGTGTAGATTAGAC 1060
 Db 217 euGlnLysilePheLysLysLeuProGlyPheGlyGluileArgValLeuGlyPheArgP 237
 QY 1061 CAAAGAAAGAAAGATGCTCAAGCTCCACAGAGATGCAACTTACGCCATCTTTAAGA 1120
 Db 237 roLysLysGluLysAspGlySerSerSerThrGluileGlnLeuMetAlailePheLysA 257
 QY 1121 GACACAGTCAGAGAGAAAGCCCTGCAAGTCACTCTCTCTCTTTGATTCCAAACAAA 1180
 Db 257 rgAspHilAlaGluAlaLysSerProAspSerHisLeuSerLeuAspSerAsnLysI 277
 QY 1181 TTGAAAGTGAAGAGTCTATCATGCAACATGAGAGAGACAGCAAGCAACCAAGATCTATC 1240
 Db 277 leGluSerGluArgileHisGlyValle---GluAspLysGlnProGluThrLysL 296
 QY 1241 TCACAGCTACAGACTCAAAAGCTGATCAGCAAGACACTAGAGAGAGACAACTCTTGG 1300
 Db 296 eutrAlaThrAspLeuLysLysLeuileleGlnLeuLeuAspGlyAspLeuSerLeuV 316
 QY 1301 ATGTGGGACAAATTCAGTTCACTGATGAATTCCTCGATCACTCCAGCCCTTGGTCTG 1360
 Db 316 alGluGlyLysileProPheGlyAspGluValThrGlyThrLeu-----PheArgProV 334
 QY 1361 ACACCAATCAAGAGTCCCAACATCTTTCTCTTATACAGAGAGATGCTACTTTGAGTC 1420
 Db 334 alThrGluProAspLeuProLysPProLeuAlaAspValThrGluLysAlaThrLeuSerP 354
 QY 1421 CAGACTTCTCTCTGTTGAAACCCAGCTTCCAGAGTGGAGCGAGAGAGCATGCTCTAC 1480
 Db 354 roGluLeuProPheValGluProArgLeuGluAlaValAspArgGluGlySerGluLeuP 374
 QY 1481 CT-----GACACTTCTGCTCCACCTGCTATGGCTTACTCTCTCTCTCTCT 1525
 Db 374 roGlyMetSerSerLysAspSerSerTrpSerProValSerAlaSerileSerArgS 394
 QY 1526 CAGAGTCCCACTTCTTTTATGATCAAGCATCTTCTCTGACTGATCAAGGCACCA 1585
 Db 394 exGluAsnLeuProSerPhe---ThrProSerilePheSerLeuAspAlaGlnSerProp 413
 QY 1586 CAGATACATGGCCCTCAGCAGACATCTAGTACAGGGGCTACCATCCCAACAGT 1645
 Db 413 roProLeuMetThrThrGlyProThrAlaLeuileProLysProThrLeuProThrileA 433
 QY 1646 ATTATTCGCAATCAGCAACTGCTCGGAATTCACATCCACCTGCTCAGATG 1705
 Db 433 epTyrSerThrileArgGlnLeuProLeuGluSerHisTrpProAlaSerSerSera 453
 QY 1706 ACACCCATCAAGTCCAGGTGGCAAGATATGTCAGACACATCTAGTAAATGATCTGT 1765
 Db 453 spArgGluLeuileThrSerSerHisAspThrileArgAspLeuAspGlyMetAspValS 473
 QY 1766 CTGACACTCTCCCTCTGAGTACCGAGAGTCCAGCAATATGTTCTGCTCCAGATC 1825
 Db 473 erAspThrProAlaLeuSerGluileSerGluLeuSerGlyTyrAspSerAlaSerGlyC 493

Db 211 --ValAsnGluThrLysThrProValLysGluLeuGlyThrAsnThrValProGluLeuP 230
QY 884 TGGAGGAGGAGGGTGGAGCTCCTCTCTGTGTAAACAGAGGTTCAAGCGCAGAGC 943
Db 230 roAlaGluGlnMetValGluPheSerValThrLeuThrAspGlnGlyThrAlaGluL 250
QY 944 TCCTGACTCCCGAGTCCCATATTACAGGAGCTACAGGAAGTCCCACTTCAGATGC 1003
Db 250 euSerAspProAsnSerProGlnThrArgGlnLeuAlaAlaLysPheGlnLeuGlnMetG 270
QY 1004 AAAAGATTTTAAAGAACTTCCAGGATTCAAAAAATCCATGTGTAGGATTTAGACAA 1063
Db 270 InLysIlePheGluLysLeuProGlyPheLysGlnIleHisValLeuGlyPheLysGlnL 290
QY 1064 AGAAGAAAAGATGGCTCAAGCTCCAGAGATGCAACTTACGGCCATCTTTAAGAGAC 1123
Db 290 YLysGluLysAspGlySerSerSerThrIleAlaArgTyrMetValAsnPheGluArg 310
QY 1124 ACAGTGCAGAAAGAAAAGCCCTGCAAGTGACTCTCTGTCCTTTGATTCACAAATG 1183
Db 310 spGlySerGluLysSerThrAlaAspAspIleSerThrIleGlySerAsnLysValG 330
QY 1184 AAAGTGAAGAGTCTATCATGGAACCATGAGGAGGACAAAGCAACCAAGAAATCTATCTCA 1243
Db 330 LuAsnGluLysValProLeuSerAlaLysGlnGluArgGluLysSerAlaThrLysLeuT 350
QY 1244 CAGCTACAGACCTCAAAAGCTGATCAGCAACACTAGAGAAAGCAACTTTG----- 1299
Db 350 hrValThrAspLeuGlnGlnLeuValAlaThrAlaLeuHisGluAspArgSerLeuProV 370
QY 1300 --GATGGGGACAATTTCAGTTCACTCATGAA----- 1329
Db 370 alAspLeuGlyThrLeuArgPheThrAspGluProIleLysProSerSerAspPheAspA 390
QY 1330 -----ATTGCTGGA-----TCACTGCCAGCCTTTGGTCTGAC-----ACCC 1366
Db 390 snAspIleGlnGlyMetValThrIleProLeuAlaGlyProAspLeuAspAspThrIleS 410
QY 1367 AATCAGAGCTGCCCATCTTT----- 1389
Db 410 erAlaGluLeuProLeuValTyrProSerProIleThrValAspGlnThrArgAspIleP 430
QY 1389 ----- 1389
Db 430 heValAspGluPheThrThrGlyIleThrAspLeuSerArgGluIleGlyGlyProGluA 450
QY 1389 ----- 1389
Db 450 spPheAspSerAsnPheIleThrSerGluProAlaPheProThrLysProSerArgGluP 470
QY 1389 ----- 1389
Db 470 roProHisAspArgSerProAspThrGluAspIleThrAspTyrGlnArgPheThrV 490
QY 1380 -----GCTGTATATACAGAGATGCTACTTTGAGTCCAGAA-----CTTC 1429
Db 490 alProPheSerAlaLeuValSerThrAspSerProAlaLysProGluAspSerTyrLeuP 510
QY 1430 CTCTGTTGAACCCAGCTTGAGAGTGGAGGAGCAGCATGGTCTACTGACACTT 1489
Db 510 roProProAlaAspGluSerAspSerAsnAspLeuIleThrAspGluSerProThrGluG 530
QY 1490 CTGGTCTCCACTCTGAGCTCTACCTCCCTGTCAGAGAGCTCCACTTTCTTTATGG 1549
Db 530 InValIleThrProAlaValTyrThrThrGlySerPheThrLeuProThrPheLeuGlnA 550
QY 1550 CATCAAGC-----A 1558
Db 550 laThrAspLysAspThrGluAlaGluMetLysLysGluLeuValGlyValThrGluProL 570
QY 1559 TCTTCTCTGACTGATCA----- 1578
Db 570 eupHelysGluAlaAspArgAspSerLeuSerGlyGlnAlaValLysMetMetAspGluL 590

QY 1578 ----- 1578
Db 590 euGluSerSerGlyAspAspIleLeuValThrThrSerThrTyrLysThrLeuProPheL 610
QY 1579 -----GGCACCCACAGATCAATGGCCACT----- 1602
Db 610 euIleGlySerSerLeuPheAlaThrGlnProGluValThrPheAlaAlaLeuP 630
QY 1603 -----GACCAGACATGCTAGTACACGGGCTCACCATCCACCAGTGAATTTCGCAA 1657
Db 630 roProAspGlnThrLeuLeu---ProThrValThrSerProPheTyrSerHisSerValV 649
QY 1658 TCAGCCAACTGGCTCTCGGAATTTCCATCCACCT----- 1692
Db 649 al-----IleAspGlnSerProGluValProAspThrLeuMetProA 663
QY 1693 -----GCATCTTCAGATGACAGCCGATCAAGTGCAGGTGGCGAAGATATGTCAGACAC 1747
Db 663 laAlaAlaSerAlaLeuProAspArgAlaSerThrGlyValGlnAspIleAlaGluL 683
QY 1748 TAGATCAAAATGATCTGTCTGACATCTCTGCCCATCTGAGTACAGAGCTCAGGAAAT 1807
Db 683 euAspGlyAlaGlyValLysSerThrAlaValLeuAspGluAlaGluHisGlySerGlyT 703
QY 1808 ATGTTTCTCTCCAGATCATTTCTTGGAGGATACCATCTCTCTCTCAGCTTTTACAGTATA 1867
Db 703 YrIleSerValGlnThrThrGluProAlaGluValThrGlnAlaProThrLeuLysYrV 723
QY 1868 TCACCATAGTCTTATGACATTTGCCCAAGGGCGAGAGCTGTAGTGTCTTCTTCAGTC 1927
Db 723 alThrSerSerMetThrThrAlaAlaLysGlyLysGluLeuValValPhePheSerL 743
QY 1928 TGGGTGTTGCTAACATGGCTTCTCCACAGACCTTTCAAAGAGCTCTCTGAGATACC 1987
Db 743 euArgValThrAsnMetHisPheSerAspAspLeuPheAsnArgSerSerGlnGluYrL 763
QY 1988 GAGCTCTGGAGCAACAATTCACACAGCTGTGTCTCCATATCTACGATCCAACTTCACAG 2047
Db 763 ysAlaLeuGluGlnGlnPheMetGlnLeuLeuLeuProTyrLeuGlnSerAsnLeuThrG 783
QY 2048 GATTAGCAACTTGAATTAATTAATTCAGAAACGGAGTGTGATTTGTAATAGCAAAA 2107
Db 783 lypHelysGlnLeuGluIleLeuAsnPheArgAsnGlySerValIleValAsnSerLysM 803
QY 2108 TCAAGTTTGTAAAGTCTGTGCGGTATTAACCTCACCAAGCTGTGCACGGGTCTTGGAGG 2167
Db 803 eLysPheAlaArgThrValProTyrAsnIleThrGluAlaValHisCysValLeuGluA 823
QY 2168 ATTTTCTGTTCTGTCAGCCCAACAACCTCCATCTGGAATATAGACAGCTACTCTCTCAACA 2227
Db 823 epPheCysAspAlaAlaAlaGlnHisLeuAsnLeuGluIleAspSerTyrSerLeuAspI 843
QY 2228 TTGACACAGCTGATCAGCAGATCCCTGCAAGTCTCTGGCTCGCGCAATTTGCCCAAT 2287
Db 843 leGluProAlaAspGlnAlaAspProCysLysPheMetAlaCysAspGluPheSerLysC 863
QY 2288 GTGTAAGAAGCAAGCAAGCTCAGGAGCGGAGTGTGCTGCAACACAGCATATGACGCC 2347
Db 863 ysIleMetAsnGluTrpThrLysGluAlaAspCysLeuCysLysProGlyTyrAlaSerG 883
QY 2348 AGGGG-----AGCCTGGAGCGTCTGGAACACAGCCCTGTGTGGCCCTGGCA 2392
Db 883 InAspGlyLeuProCysArgSerLeuCysGluMetGluProHisLeuCysAspAsnGlyG 903
QY 2393 CAAGAGAAATGCGAGTCTCTCCAGGAAGGAGGTCCATGTCAGGTTGCCAGAT 2445
Db 903 lYlYs---CysGluLeuValProGlyArgGlyAlaValCysArgSerProAsp 919
RESULT 8
Q9BWZ1
ID Q9BWZ1 PRELIMINARY; PKT; 288 AA.
AC Q9BWZ1;

```

Alignment Scores:
      Q. No.:          9.24e-59           Length:       1241
      Score:         923.00             Matches:        316
      Percent Similarity: 39.15%         Conservative:   162
      Best Local Similarity: 25.88%      Mismatches:    312
      Query Match:     15.71%            Indels:        431
      DB:              4                 Gaps:         38

US-10-007-270-1 (1-3330) x Q9UKK5 (1-1241)
      yY      170 ATTTTTCCTCCAACTGACGAAGACC---AAGATATATCTCATTACATATATACCATTCGTAA 236
      bb      16 IlePheValIleGlucLysAspPheProSerLeuThrAlaGlnThrTyrlSer --- 34
      yY      227 ACTAAGAAGCACTATAGACAATCCC-----CCAAGAATAATGAACA 262
      bb      35 IleGludIlelledGluProLysSerAlaValserPheIleLeuProgluglusSerThr 54

```


QY 308 ATATTGATTTGGCAAGCATCGAACAAGATCCGCAATTTTCCCAAGCGGGTTAAA 367
Db 75 TrpLeu-----lleArgArgArgSerileLeuPheProAsnGlyValys 90
QY 368 GTCCTGCACAGGAATCCATGAACAGATTTTAGACAGTCTTCAAGCTTATTATAGATTG 427
Db 91 IleCysProAspGluSerValaIaGluAlaValaIaAsnHisValLysTyrPheLysVal 110
QY 428 AGAGTGTCTCAGGAAGCAGATATGGGAACATATCGGATCTTCTGGATCGCATCCCTGCAC 487
Db 111 ArgValCysGlnGluAlaValTrpGluAlaPheArgThrPheTrpAspArgLeuProGly 130
QY 488 ACAGGGAATATCAGGACTGGCTCAGCATCTGCCAGCAGAGACCTTCCCTCTTTGAC 547
Db 131 ArgGluGluTyrHisTyrTrpMetAsnLeuCysGluAspGlyValThrSerIlePheGlu 150
QY 548 ATGGAAAAACTTCAGCAATTCACAGCAATCTACATTTCAAAGACTTGGCAGTATCTAA 727
Db 151 MetGlyThrAsnPheSerGluSerValGluHisArgSerLeuIleMetLysLysLeuThr 170
QY 608 CAGAGAATCTTCCCTCAGCAAGAAAGATGAATATCTGCAGAGAAGACATTTGGGAGAGCCT 667
Db 171 Tyr-----AlaLysGluThrValSerSer-----SerGlu-Le 181
QY 668 GGTGAACCAATTTGTCAATTCACAGCAATCTACATTTCAAAGACTTGGCAGTATCTAA 727
Db 181 userSerProValproValGlyAspThrSerThr-LeuGlyAspThrThrLeuSerValp 201
QY 728 GAAACCCCTCAGAAAGACAAATTCAGATGTTGCCAAGCTCTCACTTGGCGCTTCCCTC 787
Db 201 roHisProGluValAspAlaTyrGluGlyAlaSerGluSerLeuGlu----- 217
QY 788 TCACCTCTGTGATGACACCCCTCCTCAATGAAATTTCTGATAATACACCAAGCAAGA 847
Db 218 --ArgProGluGluSerIleSerAsnGluile---GluAsnValIleGluGluAlaThrL 236
QY 848 TGCTCTACACAAAGAGAAACAGAAATTCGCTGTGTTCGAGCAGCAGGGTGGAGCTCA 907
Db 236 yepProAlaGlyGluGlnIleAlaGluPhe-----S 246
QY 908 GCGTCTCTGTGTAACACAGAAAGTTCAAGCAGAGCTCGCTAGCTCCCACTCCCAATATT 967
Db 246 erIleHisLeuLeuGlyLysGlnTyrArgGluGluLeuGlnAspSerSerPheHisH 266
QY 968 ACCAGAGCTAGCAGGAAGTCCCACTTCAGATCCAAAGATATTTAAGAACTCCAG 1027
Db 266 iSGlnHisLeuGluGluGluPheIleSerGluValGluAsnAlaPheThrGlyLeuProG 286
QY 1028 GATTCAAAAAATCCATGTGTGTAGATTAGACCAAGAAAGAAAGATGCTCAAGCT 1087
Db 286 lyTyrLysGluIleArgValLeuGluPheArgSerProLysGluAsnAsp-----SerG 304
QY 1088 CCACAGAGATGCACTACGGCCATCTTTAGAGACACAGTCAGAGCAAAAGCCCTG 1147
Db 304 lyValAspValTyrTyrAlaValThrPhe-----AsnGlyGluAlaIleSerAsnT 321
QY 1148 CAGCT--GACCTCTGTCTTTGATTCCAAAGATTTGAAGTGAAGTCTATCATG 1204
Db 321 hrThrTrpAspLeuIleSerLeuHisSerAsnLysValGluAsn-----HisG 337
QY 1205 GAACATGGAGGAGCAGCAACCAACAGAAATCTATCTCACAGCTACAGCTCAAGAGC 1264
Db 337 lyLeuValGluLeuAspAspLysProThrValValTyrThrIleSerAsnPheArgAspT 357
QY 1265 TGATCAGCAAGCACTAGAGAA----- 1287
Db 357 yrlleAlaGluThrLeuGlnAsnPheLeuLeuGlyAsnSerSerLeuAsnProAspP 377
QY 1288 -----GAAACATCTTTGGATGTG- 1305
Db 377 roAspSerLeuGlnLeuIleAsnValArgGlyValLeuArgHisGlnThrGluAspLeuV 397
QY 1306 -----GGGACAAATTCAGTTCACT----- 1323

Db 397 alTrpAsnThrGlnSerSerSerleuGlnAlaThrProSerSerSerleuLeuAspAsnThrP 417
QY 1324 -----GATGAATTTGGATCACTGCCA----- 1347
Db 417 heGlnAlaAlaTrpProSerAlaAspGluSerIleThrSerSerIleProLeuAspP 437
QY 1348 -----GCCCTTGGTCTGCAGACCCCAATCAGAGTGCC----- 1380
Db 437 heSerSerGlyProProSerAlaThrGlyArgGluLeuTrpSerGluSerProLeuGlyA 457
QY 1380 ----- 1380
Db 457 spLeuValSerThrHisLysLeuAlaPheProSerLysMetGlyLeuSerSerSerProG 477
QY 1381 -----ACAT 1384
Db 477 luValLeuGluValSerSerLeuThrLeuHisSerValThrProAlaValLeuGlnThrG 497
QY 1385 CTTTGTCTGTTATAACAGAGATGCTACTTTGAGTCCAGAACTTCTCTCTGTTGAACCCC 1444
Db 497 lyLeuProValAlaSerGluGluArgThrSerGlySerHisLeu-----ValGluAspG 515
QY 1445 AGCTTGACAGAGTGCAGCGAGCAGAG-----CATGCTCTACTGACACTT 1489
Db 515 lyLeuAlaAsnValGluGluSerGluAspPheLeuSerIleAspSerLeuProSerSerS 535
QY 1490 CTTGGTCTCCACCTGCTATGGCTCTACC-----TCCC 1522
Db 535 erPheThrGlnProValProLysGluThrIleProSerMetGluAspSerAspValSerL 555
QY 1523 TGTCAGAAAGTCCA-----CCTTCTTTATGGCATTCAGCATCTCTT 1564
Db 555 euThrSerSerProTyrLeuThrSerSerIleProPheGlyLeuAspSerLeuThrSerL 575
QY 1565 CTCTCAGTGTATCAAGCACCACAGATCAATGGCCACTCACCAGCAATG----- 1614
Db 575 ysValLysAspGlnLeuLysValSerProPheLeuProAspAlaSerMetGluLysGluL 595
QY 1615 -----CTAGTACAGGGCTCACCATCC 1636
Db 595 euIlePheAspGlyGlyLeuGlySerGlySerGlyGlnLysValAspLeuIleThrTrp 615
QY 1637 CCACAGTGTATTTCTGCAATCAGCACTGGCTCGGAATTTTCATCATCTGAT 1696
Db 615 roTrpSerGluThrSerSer--GluLysSerAlaGluProLeuSerLysProTrpLeuG 634
QY 1697 CTTCAGATGACAGCCGATCAAGTCCAGGTGGCGAAGATATGTCACAGACCTA---GATG 1753
Db 634 luAspAspAspSerLeuLeuProAlaGluIleGluAspLysLeuValLeuValAspL 654
QY 1754 AAATGGATCTGCTGAC-----A 1771
Db 654 ysMetAspSerThrAspGlnIleSerLysHisSerLysTyrGluHisAspAspArgSerT 674
QY 1772 CTTCTGCCCATCTCAGGTACCA----- 1794
Db 674 hrHisPheProGluGluGluProLeuSerGlyProAlaValProIlePheAlaAspThrA 694
QY 1795 --GAGCTCAGCGAATATTTTGTGTCAGATCAFTTTCTTGAG----- 1836
Db 694 laAlaGluSerAlaSerLeuThrLeuProLysHisIleSerGluValProGlyValAspA 714
QY 1837 -----GATCAGCTCTGCTCAGCT----- 1857
Db 714 spCysSerValThrLysAlaProLeuIleLeuThrSerValaIleSerAlaSerThrA 734
QY 1858 -----TTACAGTATATCACCACTAGTT 1879
Db 734 splySerAspGlnAlaAspAlaIleLeuArgGluAspMetGluGlnIleThrGluSerS 754
QY 1880 CT-----A 1882

Db 754 erAsnTyrGluTrpPheAspSerGluValSerMetValIysProAspMetGlnThrLeuT 774
 QY 1883 TGACCAATTCGCCCAAGGCCGAGAG----- 1908
 Db 774 rPThrIleLeuProGluSerGluArgValTrpThrArgThrSerSerLeuGluLysLeuS 794
 QY 1908 ----- 1908
 Db 794 erArgAspIleLeuAlaSerThrProGlnSerAlaAspArgLeuTrpLeuSerValThrG 814
 QY 1908 ----- 1908
 Db 814 lnSerThrIysLeuProProThrThrIleSerThrLeuLeuGluAspGluValIleMetG 834
 QY 1908 ----- 1908
 Db 834 lyValGlnAspIleSerLeuGluLeuAspArgIleGlyThrAspTyrTrpGlnProGluG 854
 QY 1908 ----- 1908
 Db 854 lnValGlnGluGlnAsnGlyLysValGlySerTyrValGluMetSerThrSerValHisS 874
 QY 1908 ----- 1908
 Db 874 erThrGluMetValSerValAlaTrpProThrGluGlyAspAspLeuSerTyrThrG 894
 QY 1909 -----CTGGTAGTTCCTTCAGTCTGGTGTCTGCTAACATGGCTCTCTCA 1954
 Db 894 lnThrSerGlyAlaLeuValPhePheSerLeuArgValThrAsnMetMetPheSerG 914
 QY 1955 ACGACCTGTTCAACAAGAGCTCTGGAGTACCGAGCTCTGGAGCAACAATTCACACAGC 2014
 Db 914 luAspLeuPheAsnLysAsnSerLeuGluTyrIysAlaLeuGluGlnArgPheLeuGluL 934
 QY 2015 TGCTGGTTCCTATCTACGATCCAACTTACAGATTAAAGCACTTGAATTAATACTTAACT 2074
 Db 934 euLeuValProTyrLeuGlnSerAsnLeuThrGlyPheGlnAsnLeuGluIleLeuAsn 954
 QY 2075 TCAGAAACGGGAGTGATTGTGAATAGCAAAATGAGTTTGCTTAAGTCTGCGCGTATA 2134
 Db 954 heArgAsnGlySerIleValValAsnSerArgMetIysPheAlaAsnSerValProPro 974
 QY 2135 ACCTCACAAGGCTGTGCGCGGCTCTGGAGATTTCGTTCTGCTGCGAGCCCAACAC 2194
 Db 974 snValAsnAsnAlaValTyrMetIleLeuGluAspPheCysThrAlaTyrAsnThrM 994
 QY 2195 TCATCTGGAAATAGACACTTCTCTCAACATTGAACCAAGCTGATCAAGCAGATCCCT 2254
 Db 994 etAsnLeuAlaIleAspLysTyrSerLeuAspValGluSerGlyAspGluAlaAsnProC 1014
 QY 2255 GCAAGTCTCTGCGCTGCGCGCAATTCGCCAATGTGTAAGCAACGACGAGCTGAGGAG 2314
 Db 1014 yelLysPheGlnAlaCysAsnGluPheSerGluCysLeuValAsnProTrpSerGlyGluA 1034
 QY 2315 CGAGTGTGCTGCAAAACGAGATATGACAGCCAG-----GGGAGCTCTGG 2359
 Db 1034 laLysCysArgCysPheProGlyTyrLeuSerValGluGluArgProCysGlnSerLeuC 1054
 QY 2360 ACGTCTGGAACAGCGCTCTGTGGCCCTGGCCNCAAGAAATGGAGGCTCTCCAGGAA 2419
 Db 1054 yAspLeuGlnProAspPheCysLeuAsnAspGlyLys--CysAspIleMetProGlyH 1073
 QY 2420 AGGAGTCTCCATCAGTGTGCCAGATC-----ACTCTGAAATCAGCATACA 2467
 Db 1073 igGlyAlaIleCysArg-CysArgValGlyGluAsnTrpTrpTyrArgGlyLysCys 1092
 QY 2468 AAATAGTGTAAAAAGTTCCAAATCAACAAATAACAGGTAATCACTAAAGAAAT 2527
 Db 1093 GluGluPheValSerGluProValIleIleGlyIleThrIleAlaSerValGlyLeu 1112
 QY 2528 CTGAATTACTACCGCTAGATATGAAGATTTAACCATCAGATTGGGAAGAAATTA 2587
 Db 1113 Leu----- 1113

QY 2588 AACTGAAATGTACAAATTATCACTTAGCTATCTCAAGAGAGATGATTGCTTCTCAAG 2647
 Db 1114 -----ValIlePheSerAlaIleIleTyrPhePheIleArgThrLeuGln 1128
 QY 2648 GAAATGTGAGAGACAGCAT-----ATTCATGGTCAATCAAAATCCAGACAT 2692
 Db 1129 AlaHiHisAspArgSerGluArgGluSerProPheSerGlySerSerArgGlnProAsp 1148
 QY 2693 ACAGTCAACACT--GAGAAAT-----CAGCACACACATATTTCAATATAGAGA 2740
 Db 1149 SerLeuSerSerIleGluAsnAlaValLysTyrAsnProValTyrGluSerHisArg 1167
 RESULT 11
 Q9BZV3
 ID Q9BZV3 PRELIMINARY; PRT; 1241 AA.
 AC Q9BZV3;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Interphotoreceptor matrix proteoglycan 200.
 GN IMPG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20391216; PubMed=10542133;
 RA Kuehn M.H., Hageman G.S.;
 RT "Molecular characterization and genomic mapping of human IPM 200, a
 second member of a novel family of proteoglycans.";
 RL Mol. Cell Biol. Res. Commun. 2:103-110(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kuehn M.H., Hageman G.S.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF271379; AAG49889.1; JOINED.
 DR EMBL; AF271363; AAG49889.1; JOINED.
 DR EMBL; AF271364; AAG49889.1; JOINED.
 DR EMBL; AF271365; AAG49889.1; JOINED.
 DR EMBL; AF271366; AAG49889.1; JOINED.
 DR EMBL; AF271367; AAG49889.1; JOINED.
 DR EMBL; AF271368; AAG49889.1; JOINED.
 DR EMBL; AF271369; AAG49889.1; JOINED.
 DR EMBL; AF271370; AAG49889.1; JOINED.
 DR EMBL; AF271371; AAG49889.1; JOINED.
 DR EMBL; AF271372; AAG49889.1; JOINED.
 DR EMBL; AF271373; AAG49889.1; JOINED.
 DR EMBL; AF271374; AAG49889.1; JOINED.
 DR EMBL; AF271375; AAG49889.1; JOINED.
 DR EMBL; AF271376; AAG49889.1; JOINED.
 DR EMBL; AF271377; AAG49889.1; JOINED.
 DR EMBL; AF271378; AAG49889.1; JOINED.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF01390; SEA; 2.
 DR SMART; SM00200; SEA; 2.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS50024; SEA; 1.
 KW Receptor.
 SQ SEQUENCE 1241 AA; 138628 MW; 8362085134E6B2B CRC64;
 Alignment Scores:
 Pred. No.: 3,01e-58 Length: 1241
 Score: 915.00 Matches: 315
 Percent Similarity: 39.07% Conservative: 162
 Best Local Similarity: 25.80% Mismatches: 313
 Query Match: 15.59% Indels: 431
 DB: 4 Gaps: 38
 US-10-007-270-1 (1-3330) x Q9BZV3 (1-1241)

[illegible]

| | | | |
|----|------|---|------|
| QY | 1205 | GAACCATGGAGGAGCAGACAGCAACAGAAATCTATCTCACAGCTACAGACTCTCAAAAGCC | 1261 |
| Db | 337 | lyleuValGluLeuAspAspGlyProThrValValTyrThrIleSerAsnPheArgAspT | 357 |
| QY | 1265 | TGATCAGCAAAAGCACTAGAGAA | 1287 |
| Db | 357 | yrIleAlaGluThrLeuGlnGlnAsnPheLeuLeuGlyAsnSerSerLeuAsnProAspP | 377 |
| QY | 1288 | -----GAACAATCTTTGGATGG- | 1305 |
| Db | 377 | roAspSerLeuGlnLeuIleAsnValArgGlyValLeuArgHisGlnThrGluAspLeuV | 397 |
| QY | 1306 | -----GGGCAATTCAGTTCACT- | 1323 |
| Db | 397 | alTrpAsnThrGlnSerSerSerLeuGlnAlaThrProSerSerIleLeuAspAsnThrP | 417 |
| QY | 1324 | -----GATGAAATTCCTGGATCACTGCCA- | 1347 |
| Db | 417 | heGlnAlaAlaTrpProSerAlaAspGluSerIleThrSerSerIleProProLeuAspP | 437 |
| QY | 1348 | -----GCCTTTGGTCCTGACACCAATCAGAGCTGCC- | 1380 |
| Db | 437 | heSerSerGlyProProSerAlaThrGlyArgGluLeuTrpSerGluSerProLeuGlyA | 457 |
| QY | 1380 | ----- | 1380 |
| Db | 457 | spLeuValSerThrHisLysLeuAlaPheProSerLysMetGlyLeuSerSerSerProG | 477 |
| QY | 1381 | -----ACAT | 1384 |
| Db | 477 | luValLeuGluValSerSerLeuThrIleuHisSerValThrProAlaValLeuGlnThrG | 497 |
| QY | 1385 | CTTTTGGCTGTTATACAGCAGATCTACTTTGAGTCCAGAACTTCCTCTGTCAACCCC | 1444 |
| Db | 497 | lyLeuProValAlaSerGluGluArgThrSerGlySerHisLeu-----ValGluAspG | 515 |
| QY | 1445 | ACCTTGAGACAGTGGACGGACGAG- | 1489 |
| Db | 515 | lyLeuAlaAsnValGluGluSerGluAspPheLeuSerIleAspSerLeuProSerSerS | 535 |
| QY | 1490 | CTTGGTCTCCACCTGCTATG- | 1522 |
| Db | 535 | erPheThrGlnProValProLysGlu**IleProSerMetGluAspSerAspValSerL | 555 |
| QY | 1523 | TGTCAGAAAGTCCTCA- | 1564 |
| Db | 555 | euThrSerProTyrIleuThrSerSerIleProPheGlyLeuAspSerLeuThrSerL | 575 |
| QY | 1565 | CTCTGACTGATCAAGGCACACAGATACAAAGCCACTGACACAGCATG- | 1614 |
| Db | 575 | ysValLysAspGlnLeuLysValSerProPheLeuProAspAlaSerMetGluLysGluL | 595 |
| QY | 1615 | -----CTAGTACAGGGCTCACCATCC | 1636 |
| Db | 595 | euIlePheAspGlyGlyLeuGlySerGlySerGlyGlnLysValAspLeuIleThrTrpP | 615 |
| QY | 1637 | CCACCAAGTATTCTCGAATCAGCCCAATCGCTCTGGGAATTTCAATCCACCTGCAT | 1696 |
| Db | 615 | roTrpSerGluThrSerSer---GluLysSerAlaGluProLeuSerLysProTrpLeuG | 634 |
| QY | 1697 | CTTCAGATCAGACCGATCAAGTCAGGTGCGCAAGAT- | 1744 |
| Db | 634 | luAspAspAspSerLeuLeuProAlaGluIleGluAspLysLysLeuValLeuValAspL | 654 |
| QY | 1745 | ACTAGATGAATGGAT- | 1761 |
| Db | 654 | ysMetAspSerThrAspGlnIleSerLysHisSerLysTyrGluHisAspAspArgSerI | 674 |
| QY | 1762 | -----CTGCTGCACATC | 1774 |
| Db | 674 | leHisPheProGluGluProLeuSerGlyProAlaValProIlePheAlaAspThrA | 694 |
| QY | 1775 | CTGCCCA------TCTGAGTACCAAGACTCAGCG | 1804 |

Db 1054 ysAspLeuGlnProAspPheCysLeuAsnAspGlyIys---CysAspIleMetProGlyH 1078

QY 2420 AGGAGCTCCATCGAGTGGCGATC-----ACTCTGAAAATCAAGCATACA 2467

Db 1073 isGlyAlaIleCysArg-CysArgValGlyGluAsnTrpTrpTyrArgGlyIysHisCys 1092

QY 2468 AAACTAGTGGTAAAGAAGTTCCTCAATCAACAAATCAACAGGTANTCACTAAAGAATT 2527

Db 1093 GluGlnPheValSerGluProValIleIleGlyIleThrIleAlaSerValValGlyLeu 1112

QY 2528 CTGAATTACTGACCGTAGAATATGAAGAATTTAACCATCAAGATTGCGAAGGAAATTAA 2587

Db 1113 Leu----- 1113

QY 2588 AACTGAAAATGTACAAATTATCACTTAGGCTACTCTCAAGAGAGATGATTGGCTTCTCAAG 2647

Db 1114 -----ValIlePheSerAlaIleIleTyrPheIleArgThrLeuGln 1128

QY 2648 GAAATCGAGACAGGCAT-----ATTCAATGGCTCATCAAAATCCAGCAT 2692

Db 1129 AlaHisAspArgSerGluArgGluSerProPheSerGlySerArgGlnProAsp 1148

QY 2693 ACATGCAACT---GAGAA-----CAGCACACACCATATTTCAAATATAGAAGA 2740

Db 1149 SerLeuSerIleGluAsnAlaValLysTyrAsnProValTyrGluSerHisArg 1167

RESULT 12

Q810Y3 PRELIMINARY; PRT; 1127 AA.

AC Q810Y3 MEDLINE=22477978; PubMed=12589770;

EX Chen Q., Lee J.W., Nishiyama K., Shadrach K.G., Rayborn M.E.,

RA Hollyfield J.G.;

RT "SPACRCAN in the interphotoreceptor matrix of the mouse retina:

RL molecular, developmental and promoter analysis.";

Exp. Eye Res. 76:1-14(2003).

DR EMBL: AY174061; RAO21221.1;--

DR InterPro; IPR006209; EGF-like.

DR Pfam; PF00008; SEA_domain.

DR Pfam; PF01390; EGF; 1.

DR SMART; SM00200; SEA; 2.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS50024; SEA; 2.

SQ SEQUENCE 1127 AA; 125812 MW; B51B3D8E69606708 CRC64;

Alignment Scores:

| Pred. No.: | Length: |
|------------------------|----------|
| Score: | 1.97e-56 |
| Percent Similarity: | 891.00 |
| Best Local Similarity: | 42.01% |
| Query Match: | 27.94% |
| DBs: | 15.17% |
| Gaps: | 35 |

US-10-007-270-1 (1-3330) x Q810Y3 (1-1127)

QY 221 TCTGAACCTAGACATAGACATCCCCCAAGAAATGAAACAACACTGAAAGTACTGAAAAA 280

Db 24 SerAlaSerGluGluArgGlnGluProMetHisAlaValSerValLeuSerProGluLys 43

QY 281 ATGTCAAAATGTCA-----ACTATGAGACGAAATTCGATTGGCGAAGCAT--- 328

Db 44 SerThrAspLeuSerLeuProThrArgLysArgGlnLeuLeuAspAlaThrGluThrGly 63

QY 329 -----CGAACAAAGATCCGCATTTTCCCAAGGGGGTAAAGTCGT 373
 Db 64 ArgArgTrrPleuLeuArgArgArgSerIleleuPheProAsnGlyValIysIleCys 83
 QY 374 CCACAGGAATCCATGAACAGATTATTAGACAGTCTTCAAGCTTATTATAGATTGAGATG 433
 Db 84 SerSerGluThrValAlaGluAlaValAlaAsnHisValLysIlePheLysAlaIaargVal 103
 QY 434 TGTCCAGGAAGCAGTATGGAGCATATCGGATCTTTCTGGATCCGATCCCTCCACACAGG 493
 Db 104 CysGlnGluAlaIleTrpGluAlaPheArgThrPheTrpAspArgLeuProGlyIaargAsp 123
 QY 494 GAATATCAGGACTGGGTTCAGCATCTGCCAGCAGGAGACCTTCTCCCTTTGACATTCGA 553
 Db 124 GluIyArgHisTrpMetAsnLeuCysGluAspGlyValThrSerValPheGluMetGly 143
 QY 554 AAAAACTTCAGCAATCCCGAGGACCTCGATCTTCTCCAGCAGAGATA----- 604
 Db 144 AlaHisPheSerGlnSerValGluHisArgAsnLeuIleMetLysLysLeuAlaIyThr 163
 QY 605 AAACAGAGAGATTTCCTCGACAGAAAGATGAAATATCTGCAGAGAGACATTCGGAGAG 664
 Db 164 ArgGluAlaGluSerSerSerCysIysAspGln---SerCysGlyProGluLeuSerPhe 182
 QY 665 CCT-----GGTGAACCATTTGTCATTTCAACAGCAATCTACATTCGAAGACTGG 715
 Db 183 ProValProIleGlyGluThrSerThrLeuThrGlyAlaVal----- 196
 QY 716 GCAGTATTCTTAAGAAACCTCCAGAGAGCAAAATTCAGAGTGTGCCACGCTCCTCTG 775
 Db 197 -----SerSerAlaSerIyPro 202
 QY 776 GGCCTTCCTCT-----CTCATCTCTGTGTGACACCTCTCTCTCAATCAATCTC 822
 Db 203 GlyLeuAlaSerGluSerSerAlaAlaSerProGlnGluSerIleSerHengluIle--- 221
 QY 823 GATAATACATCTAACGACACCAAGATCCCTACACAGAAAGAGAAACAGAAATTCCTGTG 882
 Db 222 GluAsnValThrGluIuProThrGlnProAlaAlaGluGlnIleAlaGluPhe----- 239
 QY 883 TTGGAGGAGCAGAGGTGGTGTGCTCTCTCTGTGTAACACAGAGTTCAAGCAGAG 942
 Db 240 -----SerIleGlnLeuLeuGlyIysArgIySerGluGlu 251
 QY 943 CTGCGTCACTCCAGTCCCATATTACAGAGACTAGCAGGAAGTCCCACTTCAGATG 1002
 Db 252 LeuArgAspProSerSerAlaLeuIyArgLeuLeuValGluGluPheIleSerGluVal 271
 QY 1003 CAAGAATATTAAAGAACTCCAGGATTCAAAAAATCCATGTGTAGGATTAGACCA 1062
 Db 272 GluIysAlaPheThrGlyLeuProGlyIyIysGlyIleArgValLeuGluPheArgAla 291
 QY 1063 AAGAAAGAAAGATGGTCAAGCTCCACAGAGATGCAACTTACGGCCCTCTTAAAGAGA 1122
 Db 292 ProGluGluAsnAspSerGlyIleAspValHisIyValAlaValThr----- 306
 QY 1123 CACAGTCAGAGACAAAGAGCCCTGCAAGT---GACCTCTGTCTTTTGTATCCACAAA 1179
 Db 307 PheAsnGlyGluAlaIleSerAsnThrThrTrpAspLeuIleSerIleHisSerAsnLys 326
 QY 1180 ATTCAAGTCAAGGAGTCTATCATGGAACCTGGAGGAGGACAGCAACCAATCTAT 1239
 Db 327 ValGluAsn-----HisGlyLeuValGluMetAspLysProThrAlaVal 342
 QY 1240 CTCACGCTACAGACCTCAAAAGGCTGATCAGCAAGCACTAGAGGAAGAACTCTTTG 1299
 Db 343 TyrThrIleSerAsnPheArgAspIyIleAlaGluThrLeu-----HisGlnAsnPhe 360
 QY 1300 GATGTGGGGCAATTCAGTTCACGTGAATGTGTGATCACTGCCACCTTTGGTCT 1359
 Db 361 LeuMetGlyAsnSerSerLeuAsnProAsp---ProLysSerLeuGluLeuIleAsnAla 379

QY 1360 GACACCCATCAGAGTGGCC-----ACATCTTGTCTGTATATAACA 1401
 Db 380 ValLeuGlnProAspLeuProValAlaProGluGlyArgThrSerGlySerPheIleLeu 399
 QY 1402 GAGGATGCTACTTGTGATCCAGAA----- 1425
 Db 400 GluAspGlyLeuAlaSerThrGluGluLeuGluAspThrSerIleAspGlyLeuProSer 419
 QY 1426 -----CTTCTCTCTGTGTGA----- 1440
 Db 420 SerProLeuIleGlnProValProLysGluThrValProProMetGluAspSerAspThr 439
 QY 1441 -----CCCCAGCTTCAG-----ACAGTGGAC----- 1461
 Db 440 AlaLeuLeuSerThrProHisLeuThrSerSerAlaIleGluAspLeuThrLysAspIle 459
 QY 1462 GGAGCAGAGCATGCTCTA----- 1479
 Db 460 GlyThrProSerGlyLeuGluSerLeuAlaSerAsnIleSerAspGlnLeuGluValIle 479
 QY 1480 -----CCTGACACTTCT----- 1491
 Db 480 ProTrpPheProAspThrSerValGluLysAspPheIlePheGluSerGlyLeuGlySer 499
 QY 1492 -----TGGTCTCCACCTGCTATGGCTCT 1515
 Db 500 GlySerGlyLysAspValAlaAspValIleAspTrpProTrpSerGluThrSerLeuGluLys 519
 QY 1516 ACC-----TCCCTGTCAAGACTCCA----- 1536
 Db 520 ThrThrLysProLeuSerLysSerTrpSerGluGluGlnAlaLeuLeuProThrGlu 539
 QY 1536 ----- 1536
 Db 540 GlyArgGluLysLeuHisIleAspGlyArgValAspSerThrGluGlnIleIleGluSer 559
 QY 1536 ----- 1536
 Db 560 SerGluHisArgIyArgIyAspArgProIleHisPheIleGluGluGluSerHisValArg 579
 QY 1537 -----CCTTCTTATGTCATCAGC-----ATCTTCTCT----- 1566
 Db 580 SerThrIlePheIlePheValGluSerAlaThrProProThrSerProIlePheSerLys 599
 QY 1567 -----CTGACTGATCAAGCCACACAGATACATGCGCATCGACCAATCTAGTACCA 1623
 Db 600 HisThrSerAspValProAspIleAspSerTyrSerLeuThrLysProPheLeuPro 619
 QY 1624 -----GGGCTCACCATCCCCACACAGTATTATTCGCAATCAGCCAACTCGGCTCTGGGAAT 1680
 Db 620 ValThrIleAlaIleProAlaSerThrLysThrAspGluValLeuLysGluAspMet 639
 QY 1681 TCACATCCACCTGCTCATCTTCAGAT-----GACAGCCGATCAAGTGCAGTGGCGAA 1731
 Db 640 ValHisThrGluSerSerHisLysGluLeuAspSerGluValProValSerArgPro 659
 QY 1732 GATATG----- 1737
 Db 660 AspMetGlnProValTrpThrMetLeuProGluSerAspThrValTrpThrArgThrSer 679
 QY 1738 -----GTCAGACACCTAGATGAATGGATCTCTGACACTCTCT----- 1776
 Db 680 SerLeuGlyLysLeuSerArgAspThrLeuAlaSerThrProGluSerThrAspArgLeu 699
 QY 1777 -----GCCCCATCTGAGGTACCA-----CAG 1797
 Db 700 TrpLeuLysAlaSerMetThrGlnSerThrGluLeuProSerThrThrHisSerThrGln 719
 QY 1798 CTCAGCAATATGTTTCTGTC----- 1818
 Db 720 LeuGluGluGluValIleMetAlaValGlnAspIleSerLeuGluLeuAspGlnValGly 739
 QY 1819 -----CCAGATCATTTCTTG 1833

```

Db 740 ThrAspTyrTyrGlnSerGluLeuThrGluGlnHisGlyLeuAlaAspSerTyrVal 759
QY 1834 GAGATACCACTCTGCTCAGCTTTACAGTATATCACCAGTATCTTACAGCTGAC 1893
Db 760 GluMetSerThrSerVal-----HisTyrThrGluMetProIleValAlaLeuPro 776
QY 1894 CCCAAGGCC-----CGAGAGCTGGTGTAGTCTTCTTCAT 1926
Db 777 ThrLyGlyGlyValLeuSerHisThrGlnThrAlaGlyAlaLeuValAlaPheSer 796
QY 1927 CTGGGTGGTCTAATGCTCTCCAAAGACCTGTTCAACAGAGCTCTCTGGAGTAC 1986
Db 797 LeuArgValThrAsnMetLeuPheSerGluAspLeuPheAsnLeuGluTyr 816
QY 1987 CGAGCTCTGGAGCAACAAATTCACAGCTGCTGTTCCATATCTACGATCAATCTTACA 2046
Db 817 LyAlaLeuGluGlnArgPheLeuGluLeuValProTyrLeuGlnSerAsnLeuSer 836
QY 2047 GCATTTAAGCAACTTGAATACTTACTTCAGAACGGGAGTGTGATTTGATGATACAA 2106
Db 837 GlyPheGlnAsnLeuGluIleLeuSerPheArgAsnGlySerIleValValAsnSerArg 856
QY 2107 ATGAGTTTGTAAAGTCTGTGCGGTATTAACCTCACCAGGCTGTGCAGGGCTTTGGAG 2166
Db 857 ValArgPheAlaGluSerAlaProProAsnValAsnLeuAlaMetTyrArgIleLeuGlu 876
QY 2167 GATTTTGTCTGCTGAGCCCAACACTCCATCTGAAATAGACAGCTACTCTCTCAAC 2226
Db 877 AspPheCysThrThrAlaTyrGlnThrMetAsnLeuAspIleAspLysTyrSerLeuAsp 896
QY 2227 ATTGAACCAAGCTGATCAAGCAGATCCCTGCAAGTTCCTGGCTGCGGGAATTTGCCAA 2286
Db 897 ValGluSerGlyAspGluAlaAsnProCysLysPheGlnAlaCysAsnGluPheSerGlu 916
QY 2287 TGTGTAAGAAGCAACGAGCTGAGGAGCGAGTGTGCTGCACAAACAGCATATGACAGC 2346
Db 917 CysLeuValAsnProTyrSerGlyGluAlaLysCysLysCysTyrProGlyTyrLeuSer 936
QY 2347 CAGGGG-----AGCTGACGGTCTGGAACCGGCTCTGTGGCCCTGGC 2391
Db 937 ValAspGluLeuProCysGlnSerLeuCysAspLeuGlnProAspPheCysLeuAsnAsp 956
QY 2392 ACAAGGAATGCGAGGTCCTCCAGGGAAGGAGCTCCATGCAGG 2436
Db 957 GlyLys--CysAspIleMetProGlyHisGlyAlaIleCysArg 970

RESULT 13
P70628
ID P70628 PRELIMINARY; PRT; 1239 AA.
AC P70628;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PG10.2 protein.
GN PG10.2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97038366; PubMed=8883960;
RA Wang X., Brownstein M.J., Young W.S. III;
RT "Sequence analysis of PG10.2, a gene expressed in the pineal gland and
RL the outer nuclear layer of the retina.";
RL Brain Res. Mol. Brain Res. 41:269-278(1996).
DR EMBL; U76717; AAC52891.1;
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF01390; SEA; 2.

```

```

DR SMART; SM00181; EGF; 1.
DR SMART; SM00200; SEA; 2.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS00024; SEA; 2.
SQ SEQUENCE 1239 AA; 137302 MW; 78CB5AE7A73E9B6 CRC64;

Alignment Scores:
Pred. No.: 3,78e-54 Length: 1239
Score: 860.00 Matches: 283
Percent Similarity: 37.59% Conservative: 129
Best Local Similarity: 25.82% Mismatches: 268
Query Match: 14.64% Indels: 416
DB: 11 Gaps: 34

US-10-007-270-1 (1-3330) x P70628 (1-1239)
QY 224 GAACTAAAGACATAGCAATCCCCCAAGA---AATGAAACAATGAAAGTACTGAAAAA 280
Db 43 GluSerThrAspLeuSerLeuProThrArgLysArgGlnLeuLeuAspAlaThrGlu--- 61
QY 281 ATGTACAAAATGTCAACTATGACAGCAATATTCGATTTCGCAAGCATGCAACAAAAAGA 340
Db 62 -----ThrGlyArgArg---TyrProLeu-----ArgArgArgArg 72
QY 341 TCAGCATTTTCCCAACGCGGGGTTAAAGTCTGTCCACAGGAATCCATGAAACAGATTTTA 400
Db 73 SerIleLeuPheProAsnGlyValArgIleCysProSerAspThrValAlaGluAlaVal 92
QY 401 GACAGCTTCAAGCTTATATACATTCAGAGTGTGTCAGGAGCAGTATGGGAGCATAT 460
Db 93 AlaAsnHisValLysTyrPheLysAlaArgValCysGlnGluAlaIleTyrGluAlaPhe 112
QY 461 CGGATCTTTCTGATTCGATCCCTGCACAGCGGAATATCAGCATCTGGGTGAGCATCTGC 520
Db 113 ArgThrPheTyrAspArgLeuProGlyArgGluGluTyrGlnTyrTyrMetAsnLeuCys 132
QY 521 CAGCAGAGACCTCTCGCTCTTGGACATTCGAAAAAACTTCAGCATTCCTCCAGGAGCAC 580
Db 133 GluAspGlyValThrSerValPheGluMetGlyThrGlnPheSerGlnSerValGluHis 152
QY 581 CTGGATCTTCTCCAGCAGAGATA-----AAACAGAGAGTTTCCCTGCAGAGAAA 631
Db 153 ArgHisLeuIleMetGluLysLeuThrTyrThrLysGluAlaGluSerSerCysLys 172
QY 632 GAT-----GAAATATCTGCAGAGAACACATTCGGAGAGCTGTGTGAACCC 676
Db 173 AspGlnAlaCysGlyProGluLeuSerSerProValProIleGlyLeuThrSerThrLeu 192
QY 677 ATT---GTCAATTCACAGCAATCTACATTTCAAGAGCTTGGGCGATTTCTAAGAAAC 733
Db 193 AlaGlyAlaValSerSerAlaSerTyr-----AlaSerP 213
QY 734 CCTCAGAGAGCAAAATCAAGATGTTGCAACAGCTCTCACTTGGGCTTTCCCTCTCACTC 793
Db 202 Pro-GlyAlaLaserGluArgSer-Ala-----AlaSerP 213
QY 794 CTGATGACACCTCTCTCAATGAAATCTCGATATATACATCAACAGCAACAGATCCCTA 853
Db 213 roGlnGluSerIleSerAsnGluIle-----GluAsnValThrGluGlnProT 229
QY 854 CAACAGAGAGAGAACAGATTCGCTGTGTGGGAGCAGAGGGTGGAGCTCAGGCTCT 913
Db 229 hrProPro-----AlaAlaGluGlnIleAlaGluPheSerIleG 242
QY 914 CTCTGTAAACAGAGAGTTCAGGAGCAGAGCTCGCTCACTCCAGTCCCATATTTACACAG 973
Db 242 LnLeuLeuGlyLysGlnTyrSerGluGluLeuArgAspProSerSerAlaLeuTyrArgL 262
QY 974 AGCTACGAGGAAAGTCCCACTTCAGATGCAGAAAGATATTTAAGAAATTTCCAGGATCA 1033
Db 262 euLeuValGluGluPheIleSerGluValGluLysAlaPheThrGlyLeuProGlyTyrL 282
QY 1034 AAAAATCCATGTGTTAGGATTTAGACCCAAAGAAAGAAAGATGGCTCAGCTCCACAG 1093

```

```

Db      282 ysGlyIleHisValLeuAspPheArgSerProLysGluAsnGlySerGlyIleAspValH 302
QY      1094 AGATGCAACTTACGGCCATCTTTAAGACAGACAGTGCAGAGCAAGCAAAAGCCCTGCAAGT- 1152
Db      302 isTyRAlaValThr-----PheAsnGlyGluAlaIleSerAsnThrThr 317
QY      1153 --GACTCTCTCTTTTGATTCCAAACAAATGAAGTGGAGAACTCATCATGGAACCA 1210
Db      317 rpAspLeuIleSerLeuHisSerAsnLysValGluAsn-----HisGlyLeuV 333
QY      1211 TCGAGGAGCAGCAACCAACAGAAATCTATCTCACAGTCACAGCTCAGACCTCAAAAGCGCTATCA 1270
Db      333 alGluLeuAspLysProThrAlaValTyThrIleSerAsnPheArgAspTyRileA 353
QY      1271 GCAAAGCACTAGAG----- 1284
Db      353 lagluThrLeuHisGlnAsnPheLeuMetGlyAsnSerSerLeuAsnProAspProLys 373
QY      1284 ----- 1284
Db      373 erLeuGlnLeuIleAsnValArgGlyValLeuLeuProGlnThrGluIleValTrpA 393
QY      1285 ----GAAGACAATCTTTGGATGTGGGCAATTCAGTTCACTCATCAAT- 1332
Db      393 snThrGlnSerSerLeuGlnValThrSerSerIleLeuAspAsnThrLeuGlnA 413
QY      1332 ----- 1332
Db      413 lagluTrpLeuSerAlaAspGluSerIleThrThrThrThrThrIleSerProp 433
QY      1333 -----GCTGGATCAGCCAGCCTTGTGCTCAGACCCCAATCAGAG----- 1374
Db      433 heGlyPheSerSerGlyProProSerAlaThrGlyArgGluLeuHisSerGluSerThrL 453
QY      1375 -----CTGCCACATCTTT----- 1389
Db      453 euGlyAspIleValSerThrProLysLeuAlaSerProSerLysValValLeuSerSers 473
QY      1390 -----GCTGTATATA 1399
Db      473 exProGluValLeuGlyGlySerSerLeuThrLeuHisSerValThrProAlaValLeuG 493
QY      1400 CAGAGATGCTACTTTGAGTCCAGAA----- 1425
Db      493 lnIleAspLeuProValAlaProGluGlyArgThrSerGlySerSerIleLeuGluAspA 513
QY      1425 ----- 1425
Db      513 spAsnThrGluGluSerGluAspValSerIleAspValLeuProSerSerSerLeuIleG 533
QY      1426 -----CTTCTCTCTGTGAA----- 1440
Db      533 lnProValProLysGluThrValProMetGluAspSerAspMetIleLeuLeuThrS 553
QY      1441 -----CCCCACCTTGACAGTGG 1459
Db      553 erSerProHisLeuThrSerSerValIleGluAspLeuAlaLysAspIleThrThrPros 573
QY      1460 ACGGAGCAGAGCAT-----GCTCTAC 1480
Db      573 erGlyLeuAspSerLeuAlaSerArgValSerAspLysLeuAspValSerProTrpPheP 593
QY      1481 CTGACACTTCT----- 1491
Db      593 roAspThrSerValGluLysGluPheIlePheGluSerGlyLeuGlySerGlySerGlyL 613
QY      1492 -----TGCTCCACCTGCTATGGCTTCACTCC----- 1521
Db      613 yeAsnValAspValIleAspTrpProTrpSerGluThrSerLeuGluLysThrThrGluP 633
QY      1521 ----- 1521

```

```

Db      633 roLeuSerLysSerTrpSerGluGlnAspThrLeuLeuProThrGluSerIleGluL 653
QY      1521 ----- 1521
Db      653 ysLeuHisMetTyRPhethrGluGlnMetIleGluProSerAlaHisArgTyR GlyAspG 673
QY      1521 ----- 1521
Db      673 lyProIleTyRPhethrGluGluSerHisValArgSerThrIleProIlePheAlaG 693
QY      1522 -----CTGTCAGAGCTCCA----- 1536
Db      693 luSerAlaThrGlnProThrSerLeuIleSerSerLysHisThrSerSerValProAspI 713
QY      1537 -----CCTTTCTTTATGGCATCAAGCATCTTCTCTCTGA 1570
Db      713 leAspSerTyRSerValThrLysAlaProPheLeuLeuAla-----T 727
QY      1571 CTGATCAAGGCACACAGATACAAATGGCCACTGACACAG-----ACAATGCTAGTACCAG 1624
Db      727 hrIleAlaAsnThrAlaSerThrLysGluThrAspGluValAsnThrLeuLeuLysG 747
QY      1625 GGCTCACC----- 1632
Db      747 lyMetValGlnThrGluProSerSerProLysGlyLeuAspSerLysIleSerValAlaA 767
QY      1633 -----ATCCCCACAGTATTATTTCTGCATCAGCC 1663
Db      767 rgProAspMetGlnProValTrpThrIleLeuProGluSerAspThrValTrpAlaArgT 787
QY      1664 AACTGGCTCTGGA-----ATTTCATCACTCA----- 1689
Db      787 hrSerSerLeuGlyLysLeuSerArgAspThrLeuValSerThrProGluSerAlaAspA 807
QY      1690 -----CCTGCATCTTCAGATCAGACGCGATCAAGTG 1720
Db      807 rgLeuTrpLeuLysAlaSerMetThrGlnProAlaGluLeuProProThrThrHisSerI 827
QY      1721 CAGGTGGCGAAGAT-----ATGGTCAGACACCTAGATGAATGATCTGCTGCACA 1771
Db      827 leGlnLeuLysAspGluValIleMetAlaValGlnAsnIleSerLeuGluLeuAspGlnV 847
QY      1772 CTCTCTCCCATCTGAGGTACAGAGCTCAGCGAA-----T 1807
Db      847 alGlyThrAspTyRThrGlnProGluLeuThrGlnGlnAsnGlyLysValAspSerT 867
QY      1808 ATGTTTCTGCCAGATCATTTCTTGAGGATACCACTCTCTCTCCTCAGAGCTTTACGATTA 1867
Db      867 yrValGluMetProThrHisValHisTyRThrGluMetProLeuValAla---GlnProT 886
QY      1868 TCACCACTAGTTCTATGACCATTTGCCCAAGGCGGAGAGCTGGTAGTGTCTTCTCAGTC 1927
Db      886 hrLysGlyGlyValLeuSerArgThrGlnThrAlaGlyAlaLeuValValPhePheSerL 906
QY      1928 TGGTGTGCTTCAATGGCTTCTCCACAGACCTGTTCAACAGAGCTCTCTCGAGTACC 1987
Db      906 euArgValThrAsnMetLeuPheSerGluAspLeuPheAsnLysAsnSerLeuGluTyL 926
QY      1988 GAGCTCTGGACCAACATTCACACAGCTGTGTTCCATATCTACATCAATCAATCTTACAG 2047
Db      926 ysAlaLeuGluGlnArgPheLeuGluLeuValProTyRLeuGlnSerAsnLeuSerG 946
QY      2048 GATTTAAGCAACTTGAATACTTAACTTCAGAAACGGGAGTGTGATTGTGAATAGCAAAA 2107
Db      946 lyPheGlnAsnLeuGluIleLeuAsnPheArgAsnGlySerIleValValAsnSerArgV 966
QY      2108 TGAAGTTTCTAGTCTGCGGTATTAACCTCACAGGCTGTGCACGGGCTTGGAGG 2167
Db      966 alLysPheAlaGluSerValProProAsnValAsnAsnAlaIleTyRMetIleLeuGluA 986
QY      2168 ATTTTCTGTTCTGTCAGCCCAACCACTCTGGAATATAGACAGCTACTCTCTCAACA 2227
Db      986 spPheCythrThrAlaTyRThrGlnThrMetAsnLeuAspLysPheTyRSerLeuAspV 1006

```

```

QY 2228 TTGAACAAGCTGATCAAGCAGATCCCTGCAAGTTCTCTGGCTCGCGCAATTTGCCCAAT 2287
Db :|||||
1006 alGluSerGlyAspAlaAsnProCysLysPheGlnAlaCysAsnGluPheSerGluC 1026
QY 2288 GTGTAAAGAACCAACGAGTCTGAGGAGCGAGTCTCCCTCAACCAAGCATATGACAGCC 2347
Db :|||||
1026 ySLeuValAsnProTrpSerGlyGluAlaLysCysLysCysHisProlGlyTrpLeuSerV 1046
QY 2348 AGGGAGCCTGGAC-----GGTCTGGAGCAACGAGCCTCTGTGGCCCTGGCA 2392
Db :|||||
1046 alAspGluLeuProCysGlnSerValCysAspLeuGlnProAspPheCysLeuAsnAspG 1066
QY 2393 CAAGCAATGCGAGTCTCCAGGGAAGGAGTCCCATGAGG 2436
Db :|||||
1066 lyLys---CysAspValMetProGlyHisGlyAlaLleCysarg 1079

RESULT 14
Q80XH2
ID Q80XH2 PRELIMINARY; PRT; 1243 AA.
AC Q80XH2
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SPACRAN protein.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX STRAUSBERG R.;
RL Submitted (Mar-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC048863; AAH48863.1; -
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF01390; SEA; 2.
DR SMART; SM00200; SEA; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50024; SEA; 2.
SQ SEQUENCE 1243 AA; 138100 MW; 26C647136491E4D5 CRC64;

```

Alignment Scores:

```

Pred. No.: 8.09e-54 Length: 1243
Score: 855.50 Matches: 283
Percent Similarity: 38.42% Conservative: 140
Best Local Similarity: 25.70% Mismatches: 271

```

```

Query Match: 14.56% Indels: 408
DB: 11 Gaps: 36
US-10-007-270-1 (1-3330) x Q80XH2 (1-1243)
QY 221 TCTGAATAAGACATAGACAATCCCAAGAAATGAAACAACTGAAAGTAGTACAAAA 280
Db :|||||
26 SerAlaSerGluGluArgGlnProMetHisAlaValSerValLeuSerProGluLys 45
QY 281 ATGTACAAATGTCA-----ACTATGACACGAATATTCGATTGGCAAAGCAT--- 328
Db :|||||
46 SerThrAspLeuSerLeuProThrArgLysArgGlnLeuLeuAspAlaThrGluThrGly 65
QY 329 -----CGAACAAAAAGATCCGCATTTTCCCAACGGGGTTAAAGTCTGT 373
Db :|||||
66 ArgArgTrpLeuLeuArgArgArgSerIleLeuPheProAsnGlyValLysIleCys 85
QY 374 CCACAGGATCCATGAACAGATTTTACAGCTCTTCAAGCTTTATTATAGATTGAGAGTG 433
Db :|||||
86 SerSerGluThrValAlaGluAlaValAlaAenHisValLysTyrPheLysAlaArgVal 105
QY 434 TGTCAAGACAGTATGGAGCATATCGATCTTCTCGATCGCATCCCTGACACAGGG 493
Db :|||||
106 CysGlnGluAlaIleTrpGluAlaPheArgThrPheTrpAspArgLeuProGlyArgAsp 125
QY 494 GAATATCAGGACTGGGTCTGAGCATCTCCACAGGAGGACCTTCTGCTCTTTGACATTGGA 553
Db :|||||
126 GluTyrArgHisTrpMetAsnLeuCysGluAspGlyValThrSerValPheGluMetGly 145
QY 554 AAAAATTTCAGCAATTCACAGGAGCCTCGATCTTCCACAGCAGAGAATA----- 604
Db :|||||
146 AlaHisPheSerGlnSerValGluHisArgAsnLeuIleMetLysLysLeuAlaTyrThr 165
QY 605 AAACAGAGAAGTTTCCCTGACAGAAAGATGAATATATTCACAGAGACACATTGGGAGAG 664
Db :|||||
166 ArgGlnAlaGluSerSerCysLysAspGln---SerCysGlyProGluLeuSerPhe 184
QY 665 CCT-----GGTGAACCATTTGTCATTTCAACAGCAATCTTACATTTCAAGACTTGG 715
Db :|||||
185 ProValProIleGlyGluThrSerThrLeuThrGlyAlaVal----- 198
QY 716 GCATGATTCTTAAGAAAACCTTCAGAGAGCAAAATCAAGTGTGCCAAGCTCTCACTTG 775
Db :|||||
199 -----SerSerAlaSerTyrPro 204
QY 776 GGCCTTTTCCTC-----CTCACTCTGATGACACCTCTCTCAATGAATTTCTC 822
Db :|||||
205 GlyLeuAlaSerGluSerSerAlaAlaSerProGlnGluSerIleSerAsnGluIle--- 223
QY 823 GATAATACACTCAACGACACCAAGATGCTTACACAGAAAGAGAAACAGAAATTCGCTGTG 882
Db :|||||
224 GluAsnValThrGluGluProThrGlnProAlaAlaGluGlnIleAlaGluPhe----- 241
QY 883 TTGAGGAGCAGAGGGTGGAGCTCAGCGCTCTCTGTTAAACCAAGCAAGTTCAGAGCAGAG 942
Db :|||||
242 -----SerIleGlnLeuLeuLysArgTyrSerGluGlu 253
QY 943 CTGCTGTACTCCAGTCCCACTATATACAGGAGCTACAGGAGAAAGTCCCAACTTCAGATG 1002
Db :|||||
254 LeuArgAspProSerSerAlaLeuTyrArgLeuLeuValGluGluPheIleSerGluVal 273
QY 1003 CAAAGATATTAAAGAACTTCAGGATTCAGAAATCAAAAAATTCATGTGTATTAGGATTTAGACA 1062
Db :|||||
274 GluLysAlaPheThrGlyLeuProGlyTyrIleGlyIleArgValLeuLeuPheArgAla 293
QY 1063 AAGAAGAAAAGATGCTCAGCTCCAGCTCCAGAGATCAACTACGGCCATCTTTAAGAGA 1122
Db :|||||
294 ProGluGluAsnAspSerGlyIleAspValHisTyrAlaValThr----- 308
QY 1123 CACAGTGCAGAGCAAGCAAAAGCCCTCGAAGT---GACTCTCTCTCTTTTGTATTCACACAAA 1179
Db :|||||
309 PheAsnGlyGluAlaIleSerAsnThrThrTrpAspLeuIleSerLeuHisSerAsnLys 328

```



```

Db      1042 TyrProGlyTyrLeuSerValAspGluLeuProCysGlnSerLeuCysAspLeuGlnPro 1061
QY      2374 GGCCTCTGTGGCCCTGGCACAAGAGAAATCGAGGTCTCCAGGAAAGGAGGTCTCATGC 2433
Db      1062 AspPheCysLeuAsnAspGlyLys---CysAspIleMetProGlyHisGlyAlaIleCys 1080
QY      2434 AGG 2436
Db      1081 Arg 1081

RESULT 15
O46616 PRELIMINARY; PRT; 185 AA.
AC O46616;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Interphotoreceptor matrix proteoglycan 150 (Fragment).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Kuehn M.H., Hageman G.S.;
RT "Characterization And Complete cDNA Sequence Of IPM 150, A Novel Human
RL Photoreceptor Cell-Associated Chondroitin-Sulfate Proteoglycan.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047491; AAC03788.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1
FT NON_TER 185
SQ SEQUENCE 185 AA; 21204 MW; 0214806B328C36C3 CRC64;

```

Alignment Scores:

```

Pred. No.:      4,77e-53      Length:      185
Score:          841.00      Matches:      166
Percent Similarity: 85.44%      Conservative: 10
Best Local Similarity: 80.58%      Mismatches: 8
Query Match:      14.31%      Indels: 22
DB:              6      Gaps: 1

```

US-10-007-270-1 (1-3330) x O46616 (1-185)

```

QY      347 TTTTTCCTCCACGGGGTTAAAGTCTCTCCACAGCAATCCATGAACAGATTTTAGACAGT 406
Db      2 PhePheProAsnGlyValLysValCysProGlnGlnSerMetLysGlnIleLeuAlaSer 21
QY      407 CTTCAAGCTTATTATAGATTGAGAGTGTGTCCAGGAGCAGTATGGGAGCATATCGGATC 466
Db      22 LeuGlnAlaTyrArgLeuArgValCysGlnGlnAlaValTyrGlnAlaTyrArgIle 41
QY      467 TTTCTGGATCGCATCCCTGACACAGGGGAATATCAGAGCTGGGTGAGCATCTCCAGCAG 526
Db      42 PheLeuAspArgIleProAspThrGlyGluTyrGlnAspTyrValSerPheCysGlnGln 61
QY      527 GAGACCTTCCTGCTTGTGACATTCGAAACCTTCAGCAATCCAGGAGCAGCTGGAT 586
Db      62 GluThrPheCysLeuPheAspIleGlyGlnAsnPheSerAsnSerGlnGlnHisLeuAsp 81
QY      587 CTTCTCCAGCAGAGATAAAACAGAGAGATTTCCCTGACAGAAAAGATGAATATCTGCA 646
Db      82 LeuLeuGlnGlnArgIleLysGlnArgSerPheProGluArgLysAspGluValSerThr 101
QY      647 GAGAGACATTCGGAGAGCCTGGTGAACCATTTCTATTCACACAGCAATCTACATTTCA 706
Db      102 GluLysThrLeuGlyGluProSerGluThrIleValValSer--Thr----- 116
QY      707 AAGACTTGGGCAGTATTCTAAGAAAACCTTCAGAGAGCAATTCAGATGTTGCCAAG 766

```

```

Db      117 -----AspValAlaSerV 121
QY      767 TCTCACTTGGGCTTTCCCTCTCTCCTCTGATGACACCCCTCTCAATGAATTTCTCGATA 826
Db      121 alSerLeuGlyProPheProValThrProAspAspThrLeuLeuAsnGluIleLeuAspA 141
QY      827 ATACACTCAACGACACCAAGATGCTTACACAGAAAGAGAAACAGAAATTTCTGCTGTG 886
Db      141 snlalaLeuAsnAspThrLysMetProThrThrGluArgGluThrGluLeuAlaValSerG 161
QY      887 AGGAGCAGAGGGTGGAGCTCAGCGTCTCTCTGTGTAACCAAGAGTTCAAGGCAGAGCTCG 946
Db      161 luGluGlnArgValGluLeuSerIleSerLeuIleAsnGlnArgPheLysIleGluLeuA 181
QY      947 CTGACTCCCAAGTCC 960
Db      181 laAspSerGlnSer 185

```

Search completed: March 4, 2004, 18:52:41
Job time : 230 secs